

MESTRADO INTEGRADO EM MEDICINA

Inteligência Artificial na deteção de Fraturas: perspetiva atual e futura

Margarida Reis Bragança

M

2025



Inteligência Artificial na detecção de Fraturas: perspetiva atual e futura

Artigo de Revisão Bibliográfica

Dissertação de candidatura ao grau de Mestre em Medicina, submetida ao Instituto de Ciências Biomédicas Abel Salazar – Universidade do Porto.

Estudante: Margarida Reis Bragança

Aluna do 6º ano do Mestrado Integrado em Medicina

Instituto de Ciências Biomédicas Abel Salazar – Universidade do Porto

Endereço de correio eletrónico: Margaridarb11@gmail.com | up201905770@up.pt

Coorientador e Orientador: Professor Doutor Pedro Filipe Ferreira Cardoso

Assistente Graduado do Serviço de Ortopedia da Unidade Local de Saúde de Santo António

Professor Auxiliar Convidado do Instituto de Ciências Biomédicas Abel Salazar - Universidade do Porto, Unidade Local de Saúde de Santo António, E.P.E. – Hospital de Magalhães Lemos

Endereço de correio eletrónico: pffc Cardoso@gmail.com

Porto, junho de 2025

Dissertação de candidatura ao grau de Mestre em Medicina, submetida ao Instituto de Ciências
Biomédicas Abel Salazar – Universidade do Porto.

Estudante



(Margarida Reis Bragança)

Sob a orientação e coorientação

Assinado por : **PEDRO FILIPE FERREIRA CARDOSO**

Num. de Identificação: 05911750

Data: 2025.06.08 19:05:09+01'00'



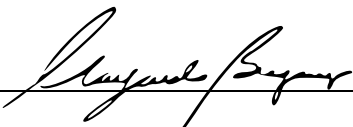
(Professor Doutor Pedro Filipe Ferreira Cardoso)

Porto, junho de 2025

Declaração de Integridade Científica

Eu, Margarida Reis Bragança, portadora do n.º de estudante 201905770, encontro-me atualmente a finalizar o ciclo de estudos no 6.º ano do Mestrado Integrado em Medicina do Instituto de Ciências Biomédicas Abel Salazar - Universidade do Porto. Declaro que esta dissertação é da minha autoria e não foi utilizada anteriormente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores respeitam as regras de referenciação, estando devidamente elencadas ao longo do texto e nas referências bibliográficas. Tenho conhecimento de que a prática de plágio e auto-plágio incorre uma infração académica.

Porto, junho de 2025.



(Margarida Reis Bragança)

Dedicatórias

À minha família por serem sempre um motor catalisador, onde não há descida que não se traduza numa oportunidade de ganhar impulsão para a subida. Os responsáveis legais por todo e qualquer sucesso que me possa ser atribuído.

Aos meus amigos, por terem sido os meus maiores pilares quando estava longe dos alicerces de casa.

A Biomédicas, onde entrei mas de onde jamais sairei.

Agradecimentos

Ao Prof. Dr^o. Pedro Cardoso, a quem manifesto o meu profundo agradecimento pela disponibilidade demonstrada ao longo da execução desta dissertação, bem como pelo valioso contributo para o rigor científico e metodológico da mesma. A sua orientação não se traduziu apenas em correções técnicas, mas também numa presença tranquila e acessível, criando um ambiente propício à reflexão crítica.

Resumo

Introdução:

A Inteligência Artificial (IA) tem marcado o seu cunho como uma mais-valia no domínio da saúde, nomeadamente na imagiologia médica. A sua aplicação na deteção de fraturas, uma das principais causas de admissão em serviços de urgência, configura uma oportunidade de otimização diagnóstica, particularmente em contextos de sobrecarga clínica e com potencial de ferramenta mitigadora de erro clínico associado. A presente revisão propõe-se a analisar o desempenho da IA na deteção de fraturas, com especial enfoque sobre as localizações anatómicas do punho, anca e, de forma isolada, no contexto pediátrico. Para além da performance intrínseca dos modelos, é igualmente avaliada a sua eficácia relativa face à avaliação médica convencional, bem como o desempenho dos sistemas atualmente disponíveis no mercado.

Métodos:

Procedeu-se a uma revisão da literatura publicada entre 2015 e dezembro de 2024, com recurso às bases de dados *PubMed*, *OrthoSearch* e *ScienceDirect*, mediante a utilização dos *Boolean operators*: “*Artificial Intelligence*”, “*Fracture Detection*”, “*Deep Learning*”, “*Convolutional Neural Networks*” e “*Medical Imaging*”. Foram incluídos estudos originais e revisões sistemáticas redigidos em língua inglesa, desde que contemplassem a aplicação de algoritmos de IA ou softwares comerciais na deteção de fraturas. A seleção privilegiou trabalhos mais recentes que apresentassem métricas de desempenho e comparações diretas com profissionais de saúde, bem como aqueles que se debruçaram sobre as regiões anatómicas do punho e anca, e aqueles que analisaram exclusivamente a faixa etária pediátrica.

Desenvolvimento:

Os modelos de IA mais amplamente analisados fundamentam-se em arquiteturas de *Convolutional Neural Networks* (CNNs), como *YOLOv5*, *EfficientNet*, *InceptionV3* e *RetinaNet*, revelando desempenhos frequentemente equiparáveis, e por vezes superiores, aos obtidos por especialistas humanos. Estratégias complementares, como a segmentação anatómica e a utilização de técnicas como *Grad-CAM*, contribuem significativamente para a interpretabilidade dos resultados. Verificou-se uma associação clara entre a qualidade, a quantidade e o grau de pré-processamento das imagens de treino e o desempenho final dos modelos. Reforça-se ainda o valor acrescido de

uma validação cruzada rigorosa da *ground truth*, com base em tomografia computadorizada (TC) ou ressonância magnética (RM). Identificaram-se, contudo, limitações ao nível da deteção de fraturas subtis e complexas, nomeadamente em estruturas em desenvolvimento, como se verifica na idade pediátrica. No que tange às soluções comerciais, modelos como o *BoneView*® demonstraram ganhos relevantes de sensibilidade e especificidade em cenários clínicos reais.

Conclusão:

A IA revela um papel preponderante na deteção de fraturas, com desempenhos que se aproximam, e por vezes superam, os obtidos por ortopedistas e radiologistas, sobretudo em contextos clínicos de elevada exigência. Persistem, todavia, desafios significativos, como a heterogeneidade metodológica, a escassez de bases de dados devidamente anotadas e a limitada padronização dos estudos existentes. Torna-se imperioso investir em validações multicêntricas, na integração estruturada da IA nos fluxos clínicos e no desenvolvimento de diretrizes normativas que orientem a aplicação rigorosa e ética destas ferramentas na prática médica.

Keywords: *Artificial Intelligence, Fracture Detection, Deep Learning, CNN, Wrist Fractures, Hip Fractures, Pediatric Imaging, Clinical Decision Support; Clinical Workflow*

Abstract

Introduction:

Artificial Intelligence (AI) has been establishing itself as a promising tool in the field of healthcare, particularly in medical imaging. Its application in fracture detection, one of the leading causes of emergency department admissions, represents an opportunity for diagnostic optimization, especially in settings of clinical overload and with potential to mitigate associated diagnostic errors. This review aims to examine the performance of AI in fracture detection, with a specific focus on anatomical locations such as the wrist, hip, and, in isolation, within the paediatric context. Beyond the intrinsic performance of AI models, the review also assesses their effectiveness relative to conventional medical evaluation, as well as the performance of currently available commercial systems.

Methods:

A literature review was conducted covering publications between 2015 and December 2024, using the databases PubMed, OrthoSearch, and ScienceDirect, through the use of Boolean operators with the keywords: *“Artificial Intelligence”*, *“Fracture Detection”*, *“Deep Learning”*, *“Convolutional Neural Networks”*, and *“Medical Imaging”*. Original studies and systematic reviews written in English, provided they addressed the application of AI algorithms or commercial software in fracture detection were included. The selection prioritized more recent studies that presented performance metrics and direct comparisons with healthcare professionals, as well as those that focused on the anatomical regions of the wrist and hip, and those that exclusively analyzed the pediatric age group.

Development:

The most extensively analysed AI models are based on Convolutional Neural Network (CNN) architectures, such as YOLOv5, EfficientNet, InceptionV3, and RetinaNet, which have demonstrated performance levels often comparable to, and at times surpassing, those of human experts. Complementary strategies such as anatomical segmentation and techniques like Grad-CAM significantly contribute to enhancing model interpretability. A clear association was observed between the quality, quantity, and level of preprocessing of training images and the final performance of the models. The added value of rigorous cross-validation of the ground truth, using Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), is also highlighted. However, limitations were noted in detecting more subtle and complex fractures, particularly in

developing anatomical structures, as seen in paediatric cases. Among commercial solutions, tools such as BoneView© have shown significant improvements in sensitivity and specificity in real-world clinical scenarios.

Conclusion:

AI demonstrates remarkable potential in fracture detection, with performance levels that rival—and in some cases exceed—those of orthopaedists and radiologists, especially in demanding clinical settings. Nonetheless, several challenges persist, including methodological heterogeneity, a lack of rigorously annotated datasets, and the limited standardization of existing studies. It is therefore imperative to invest in multicentric validation studies, in the structured integration of AI into clinical workflows, and in the development of normative guidelines to ensure the rigorous and ethical implementation of these tools in medical practice.

Keywords: *Artificial Intelligence, Fracture Detection, Deep Learning, CNN, Wrist Fractures, Hip Fractures, Pediatric Imaging, Clinical Decision Support; Clinical Workflow*

.

Lista de abreviaturas

AP-Antero-Posterior
ANN-*Artificial neural network*
AUC – *Area Under the Curve*
CAD – *Computer-Aided Diagnosis*
CE – *Conformité Européenne*
CNNs – *Convolutional Neural Networks*
DL – *Deep Learning*
DNN - *Deep Neural Networks*
FRD – Fratura do Rádio Distal
FAF- Fratura Atípica do Fémur
FDA – *Food and Drug Administration*
FPF – Fratura periprotésica do fémur
FPN - *Feature Pyramid Networks*
FRE – *Fracture of the Radial Epiphysis*
FRM – *Fracture of the Radial Metaphysis*
FUE – *Fracture of the Ulnar Epiphysis*
FUM – *Fracture of the Ulnar Metaphysis*
IA – Inteligência Artificial
IoU - *intersection over union*
ML – *Machine Learning*
MS COCO – *Microsoft Common Objects in Context*
MURA – Musculoskeletal Radiographs
PAC – *Picture Archiving and Communication System*
RM – Ressonância Magnética
ROI – *Region of Interest*
TC – Tomografia Computorizada
VCS-*Vancouver Classification System*

Índice

Agradecimentos.....	i
Resumo	ii
Abstract	iv
Lista de abreviaturas.....	vi
Índice	vii
Lista de figuras.....	viii
Lista de tabelas	x
Introdução	1
Objetivos.....	3
Metodologia	3
1. Inteligência Artificial na medicina.....	4
1.1 Background histórico	4
1.2. Funcionamento da Inteligência Artificial	4
1.3 Aprendizagem de modelos de Inteligência Artificial para a detecção de fraturas	6
2. Usos da Inteligência Artificial na Ortopedia	7
3. Inteligência Artificial na detecção de fraturas.....	8
3.1. Epidemiologia e Aplicabilidade de modelos de Inteligência Artificial	8
3.2. Inteligência Artificial na detecção de fraturas do punho	10
3.2.1. Epidemiologia das fraturas do punho.....	10
3.2.2. Performance de modelos disponíveis.....	10
3.2.3 Performance de Inteligência Artificial vs. Médico	12
3.3. Inteligência Artificial na detecção de fraturas da anca	14
3.3.1. Epidemiologia das fraturas da anca	14
3.3.2. Performance de modelos disponíveis.....	15
3.3.3. Performance de Inteligência Artificial vs. Médico	17
3.4. Inteligência Artificial na detecção de fraturas pediátricas	21
3.4.1. Epidemiologia das fraturas pediátricas.....	21
3.4.2. Performance dos modelos de IA na detecção de fraturas pediátricas.....	21
4. Modelos Comerciais	23
5. Onde se coloca a IA no futuro e entraves na utilização	26
6. Discussão	28
7. Conclusão.....	30

Apêndices	31
Referências Bibliográficas.....	47

Lista de figuras

Figura 1 - Esquematização dos subtermos de IA

Figura 2 - Processo de treino dos modelos de IA aplicados na detecção de fraturas

Figura 3 - Cálculo de IoU

Figura 4- Classificação AO adaptada em Jiménez- Sánchez et al.

Figura 5- Classificação AO/AOT

Lista de tabelas

Tabela I – Estudos de modelos de IA aplicados à detecção de fraturas da região do punho

Tabela II – Estudos de modelos de IA aplicados à detecção de fraturas da região da anca

Tabela III - Estudos de modelos de IA aplicados à detecção de fraturas pediátricas

Introdução

A definição de Inteligência Artificial (IA) remonta ao ano de 1955, com John McCarthy a conceptualizá-la como “a ciência e engenharia de fazer máquinas inteligentes”. Atualmente, é tida como a capacidade de qualquer máquina ou tecnologia recriar habilidades cognitivas humanas, tais como o planeamento de tarefas, a compreensão de linguagem, o reconhecimento de padrões e a resolução de problemas. Alberga ainda, a competência de reconhecer os seus próprios erros e corrigi-los, o que em muito se assemelha à aprendizagem experimental.^{1,2}

No campo médico, tem sido notória num variadíssimo espectro de atuação, desde o diagnóstico até ao tratamento. É possível discernir dois grandes ramos da aplicação da IA na medicina: o virtual e o físico. O ramo virtual compreende abordagens informáticas que se expandem desde a gestão e otimização dos registos eletrónicos em saúde até à cooperação direta na prática clínica, fornecendo orientações na tomada de decisões.² Por outro lado, o domínio físico compreende dispositivos médicos projetados tanto para ter um papel direto sob o doente como no auxílio do cirurgião, como ocorre nas cirurgias com recurso ao robot Da Vinci.³

No que respeita à ortopedia, tem-se verificado um aumento substancial da utilização desta tecnologia. Cabitza et al. mostrou uma tendência crescente do uso de IA em Ortopedia de quase dez vezes superior comparativamente a 2010. Demonstrou, ainda, que a IA era usada essencialmente para fins de diagnóstico como, por exemplo, a previsão ou deteção de osteoartrite.^{4,5}

As fraturas correspondem a uma entidade nosológica de frequente de ida ao serviço de urgência, tornando-se por isso relevante perceber como se afiguram num contexto social e qual a sua tendência epidemiológica. Com pico de incidência bimodal (crianças e adolescentes e na população idosa), vários fatores condicionam a sua incidência, tais como, a diabetes, idade, a obesidade ou até a farmacoterapia.^{6,7}

Torna-se pertinente destacar, que em Portugal, à semelhança do que se verifica noutros países desenvolvidos, denota-se uma tendência de envelhecimento populacional onde a população idosa regista um índice de envelhecimento de 182 (i.e. proporção de 182 idosos para cada 100 jovens), sendo o 2º valor mais alto da União Europeia.⁸ Esta propensão global de envelhecimento tem uma repercussão traduzida no facto de se esperar uma duplicação do número de fraturas osteoporóticas até 2040.⁸

Nos Estados Unidos da América, em média, são vistos nos serviços de urgência mais de 2 milhões de doentes por ano com fraturas. Representam uma percentagem de 13,9% das lesões pediátricas e 15% em adultos.⁹ A pressão que estes números representam sobre os sistemas de saúde, aliado ao cansaço médico e à ausência de radiologistas em muitos dos hospitais periféricos durante o período da noite, podem apresentar-se como um fator causal para que o período entre as 20h00 e as 2h00 — intervalo em que se observa a maior incidência desses equívocos¹⁰ A taxa de erro diagnóstico varia entre 1% e 3,7%, sendo

que apenas 33% desses casos se deve à imperceptibilidade das lesões nas imagens. Assim, a maior parte dos erros está relacionada a interpretações incorretas por parte dos profissionais de saúde.¹¹

Várias são as consequências advindas de um diagnóstico errado, desde a má consolidação com restrição da amplitude de movimento, osteoartrite pós-traumática até osteonecrose.¹² No sentido de ajudar a mitigar este mesmo erro, a IA pode apresentar-se como um auxílio para os profissionais de saúde, tendo-se verificado taxas de 50% de redução de erro aquando da combinação da sua utilização.¹³

Adicionalmente, estimou-se que a IA teria a capacidade de processar mais de 250 milhões de imagens por, aproximadamente, um milhar de euros, o que representa uma potencial poupança na ordem dos biliões de euros.¹⁴ Além do supramencionado, os ganhos de produtividade e fluxo de trabalho que derivam da interpretação de imagens assistida por IA, poderão tanto reduzir a carga laboral dos médicos no exercício clínico, como ter implicações em trabalhos administrativos, na codificação e faturação, agendamento de salas de cirurgia e consultas.^{2,15}

Objetivos

Com o intuito primordial de investigar o estado de arte da integração da Inteligência Artificial na detecção de fraturas ósseas, bem como a eficácia do que está atualmente comercializado neste sentido, a presente revisão debruça-se sobre uma análise crítica dos estudos disponíveis. Adicionalmente, pretende-se refletir sobre o potencial de integração destas ferramentas no contexto clínico real, sublinhando as suas mais-valias operacionais, bem como as limitações inerentes ao seu uso na prática e o que será profícuo pensar no futuro.

Metodologia

Foi conduzida uma revisão bibliográfica entre 2015 e dezembro de 2024 sobre a aplicabilidade da IA na detecção de fraturas, com especial enfoque em fraturas do punho, anca e pediátricas, bem como na comparação do desempenho dos modelos com o dos profissionais de saúde.

A pesquisa foi efetuada nas bases de dados *PubMed*, *OrthoSearch* e *ScienceDirect*, utilizando os seguintes *Boolean operators*, usado isoladamente ou em combinação: “*Artificial Intelligence*”, “*Fracture Detection*”, “*Deep Learning*”, “*Convolutional Neural Networks*” e “*Medical Imaging*”.

Os critérios de inclusão contemplaram artigos originais e revisões sistemáticas publicados em língua inglesa que abordassem especificamente a utilização de algoritmos de IA ou *softwares* comerciais na detecção de fraturas. A seleção foi feita tendo em conta o título, resumo e corpo. Foram especialmente valorizados os estudos que apresentassem métricas de desempenho diagnóstico (como acurácia, sensibilidade, especificidade e AUC) e/ou que comparassem os resultados da IA com os de médicos especialistas, ou ainda que explorassem a integração da IA na prática clínica. Foram privilegiados ainda os artigos mais recentes, tendo sido analisadas sistematicamente as referências bibliográficas dos mesmos com o intuito de identificar estudos adicionais relevantes não contemplados na pesquisa inicial.

1. Inteligência Artificial na medicina

1.1 *Background* histórico

Destarte do início do uso da IA na medicina datar da década de 1980, várias foram as limitações que atrasaram a sua implementação.¹⁵ Uma das aplicações pioneiras remonta a 1976, com o modelo CASNET. Concebido como uma rede causal-associativa, que articulava conhecimento específico de uma patologia com os dados clínicos individuais de cada doente. Desta forma, formulava recomendações quanto à conduta terapêutica, o que representa um advento face aos atuais sistemas de apoio à decisão clínica.¹⁶

Apenas no início do século XXI houve um crescimento exponencial da informação e recursos disponíveis, facilmente percecionado pela pesquisa simples em motores de busca como a *Pubmed*, onde surgiu um aumento significativo dos artigos disponíveis nos últimos anos.¹⁷ Com o desenvolvimento dos seus sistemas, há agora uma maior dispersão de utilização na medicina, marcando o início de uma nova era onde modelos preditivos podem ser usados no diagnóstico, tratamento, previsão de resposta terapêutica e medicina preventiva. Esta aplicabilidade tem-se traduzido em maior eficiência de fluxo de trabalho.^{18 19}

1.2. Funcionamento da Inteligência Artificial

A Inteligência Artificial articula a ciência com a engenharia, de forma a recriar os processos da inteligência humana, com recurso a sistemas não biológicos. De forma a poder executar análises, serve-se de bases de dados, através das quais desenha padrões e correlações que lhe permitem tirar inferências.²⁰ Estas bases incorporam um espectro que inclui imagens, números, textos, entre outras, havendo *a posteriori* uma seleção dos elementos mais relevantes nestas, que constituem as variáveis que serão analisadas através desses mesmos algoritmos de IA.¹⁸

Dada a extensão de conceitos que a IA integra, destacam-se 3 grandes termos de relevância major (Fig. 1) na aplicação desta na área médica: *Machine Learning* (ML); *Deep Learning* (DL) e *Convolutional Neural Network* (CNN):

ML compreende um ramo da Inteligência Artificial, que de uma forma simplista é tida como a capacidade dos algoritmos aprenderem com a experiência, melhorando a sua performance sem necessidade de ser programados continuamente para tal. No que concerne a imagens médicas possui a capacidade de detetar e classificar lesões, segmentar imagens, analisar dados e estabelecer listas de prioridade de triagem. O processo de aprendizagem pode ser, essencialmente, classificado como supervisionado ou não supervisionado.^{21,22}

Na aprendizagem supervisionada, o modelo é treinado utilizando dados previamente rotulados e

categorizados, aprendendo a tirar inferências de resultados com base nestes. Concomitantemente destaca-se por ser conhecedor dos *outputs* procurados. No contexto médico, essa abordagem tem sido amplamente aplicada tanto no diagnóstico como no prognóstico.²³ Impera, contudo, a limitação da necessidade de um volume extenso de dados corretamente rotulados e representativos, sob pena de ser executada uma generalização incorreta aquando da transposição do modelo na realidade. É ainda de ter em consideração o tempo que consome e os encargos económicos que acarreta.²⁴

Por oposição, a aprendizagem não supervisionada não requer dados rotulados. De forma autónoma, o modelo procura encontrar semelhanças e padrões entre os dados, sem que haja uma saída pré-definida ou uma categorização prévia durante o treino.²⁴ A título de exemplo, tem capacidade de agrupar doentes com sinais e sintomas idênticos em grupos, bem como agregar imagens médicas com características semelhantes entre si. Ressalva-se, contudo, a difícil aplicabilidade clínica direta por serem resultados difíceis de interpretar, requerendo validação adicional.^{20,25}

DL, por sua vez constitui uma subárea do ML, diferenciando-se deste uma vez que se distingue deste pelo modo como processa os dados. No ML tradicional, é necessário extrair manualmente um conjunto de características específicas como contornos, bordos, formas ou texturas que são depois utilizadas por um classificador multivariado para gerar a decisão final. Já no DL, a imagem completa é fornecida diretamente a uma rede neural profunda, que aprende automaticamente quais as características mais relevantes e produz diretamente uma decisão como saída, dispensando a engenharia manual de atributos.^{26,27}

Do ponto de vista computacional, o DL caracteriza-se pela utilização de redes compostas por unidades elementares interligadas. A disposição hierárquica e sequencial destas subunidades em múltiplas camadas confere ao modelo a capacidade de gerar representações progressivamente mais abstratas apesar da elevada complexidade dos dados de entrada, como é o caso da informação visual, nomeadamente imagens.^{28,29}

Para a devida compreensão da estrutura e funcionamento das arquiteturas de DL, impõe-se, inevitavelmente, a introdução do conceito de *Artificial Neural Network* (ANN). As ANNs têm origem no conceito de neurónio artificial, o qual constitui uma aproximação da estrutura computacional da célula nervosa biológica. Apesar da diversidade de modelos existentes, o funcionamento canónico de um neurónio artificial converge na atribuição de pesos numéricos aos sinais de entrada, no cálculo da respetiva soma ponderada e na aplicação subsequente de uma função matemática denominada função de ativação, a qual produz um valor de saída. Como um todo, as ANNs são formadas por uma multiplicidade destes mesmos neurónios interligados, organizados, em regra, sob a forma de camadas. Um modelo convencional de ANN é composto, de uma forma geral, por duas a três camadas.^{24,30}

Importa, por outro lado, introduzir as *Deep Neural Networks* (DNNs), que constituem uma classe

mais especializada de ANN. Entre estas, a *Convolutional Neural Network* (CNN) assume-se como a mais amplamente utilizada e investigada. Embora preserve a estrutura básica das ANNs clássicas, a CNN distingue-se por um conjunto de características computacionais singulares, sendo de destacar, entre elas, a profundidade da rede, frequentemente compreendida entre 10 e 30 camadas, podendo, em casos particulares, exceder mesmo as mil camadas.³¹

1.3 Aprendizagem de modelos de Inteligência Artificial para a deteção de fraturas

De forma a utilizar os recursos de IA acima referidos em contextos clínicos concretos, torna-se imperativo recorrer a um processo de aprendizagem (descrito Fig.2), sendo o mais amplamente utilizado a aprendizagem supervisionada, explanado anteriormente. Apesar da variabilidade de algoritmos existentes para a deteção automatizada de fraturas, denota-se que o processo de aprendizagem constitui um denominador comum entre eles, ainda que subsistam variações metodológicas, o que depois se poderá traduzir eventualmente em diferentes performances.³²

No cômputo geral, este processo de aprendizagem assenta na alteração dos pesos sinápticos dos neurónios artificiais na ponderação de dados fornecidos como input. Quando o dado de entrada é, por exemplo, uma imagem médica, esta é submetida à rede, que procede ao seu processamento interno com base nos pesos vigentes, produzindo um output preditivo.³³ Este resultado é, então, comparado com a designada *ground truth*, ou seja, a classificação atribuída no rótulo da imagem, e calcula-se o erro de predição. Este erro, por sua vez, é propagado ao longo da rede com o intuito de reajustar os pesos dos neurónios, de modo que, em futuras exposições com o mesmo exemplo, o erro seja atenuado.³⁴

No que toca à seleção de algoritmo computacional, vários estudos no âmbito do trauma ortopédico fazem uso de CNNs já consolidadas, disponíveis em bibliotecas públicas, como a *YOLOv3*, *ResNet*, *U-Net*, *RetinaNet* ou *NFNet*. Ressalva-se que algumas destas arquiteturas foram inicialmente desenvolvidas e treinadas para tarefas não médicas, e que a sua adaptação ao contexto clínico através de *transfer learning* permite reorientar a rede para reconhecer padrões específicos médicos. Esta técnica assenta na transferência de conhecimento adquirido por uma rede neural previamente treinada em domínios distintos, geralmente com grandes volumes de dados, para uma nova tarefa específica. Parte-se do pressuposto de que diferentes tarefas visuais partilham elementos estruturais comuns, como a identificação de arestas, formas simples e texturas.³⁵

No que concerne às bases de dados, neste tema em concreto, as redes são frequentemente treinadas com conjuntos de imagens previamente rotuladas, geralmente com base em pareceres clínicos emitidos por radiologistas ou ortopedistas. Estes especialistas podem apoiar-se em registos clínicos complementares, ou até em exames de Tomografia Computorizada (TC) ou Ressonância

Magnética (RM), para fundamentar a sua decisão e estabelecer a *ground truth*.³⁶ As imagens utilizadas para treino podem ainda ser fornecidas com realce das regiões anatómicas de interesse. São frequentemente extraídas de bases de dados públicas, como a MURA, que contém mais de 40.000 radiografias rotuladas, ou provenientes de sistemas hospitalares de arquivamento, como o *Picture Archiving and Communication System* (PACS).^{37,38} De forma a aumentar a diversidade do conjunto de dados e, conseqüentemente, a robustez do modelo, é comum a aplicação de técnicas de *data augmentation*, como rotação, espelhamento, alteração de escala ou variações de brilho e contraste, visando melhorar a capacidade de generalização da rede perante novos dados.^{39,40}

Durante o treino, é habitual reservar-se uma parte das imagens para validação interna contínua, permitindo monitorizar o desempenho do modelo e ajustar os seus hiperparâmetros tais como a taxa de aprendizagem e o número de épocas.⁴¹ A taxa de aprendizagem corresponde à amplitude das modificações introduzidas nos parâmetros da rede após cada iteração, sendo essencial para manter um equilíbrio entre a rapidez de convergência e a estabilidade do processo de treino. Estas avaliações contínuas, bem como volumes elevados de imagens de treino são muito relevantes no sentido de evitar o fenómeno de *overfitting*, que decorre de uma internalização em demasia das particularidades do conjunto de dados de treino, ao invés de inferir padrões gerais que caracterizam as classes, comprometendo a capacidade de generalização do modelo e conseqüente redução do desempenho aquando da confrontação com novos dados.⁴²

Por fim, um conjunto distinto de imagens é reservado para a fase de teste, contendo dados nunca antes apresentados à rede durante o treino. Este conjunto é determinante para avaliar a real capacidade do modelo em generalizar o conhecimento adquirido e, assim, aferir a sua aplicabilidade em cenários clínicos reais. É precisamente com base neste conjunto de teste que se realizam as avaliações estatísticas de desempenho, podendo os estudos optar por comparar os resultados com decisões clínicas, ou até explorar abordagens de colaboração entre médicos e algoritmos.⁴³ Adicionalmente, nesta fase, alguns autores recorrem a estratégias de *Ensemble Machine Learning*, que combinam *outputs* de diferentes modelos para culminar numa melhor performance.⁴⁴

2. Usos da Inteligência Artificial na Ortopedia

A IA tem ampla aplicabilidade na área do trauma ortopédico, como ferramenta de auxílio diagnóstico, planeamento terapêutico com suporte cirúrgico e prognóstico. Um dos primeiros estudos piloto na área diagnóstica, foi desenvolvido por Brett et al. com um modelo que identificava fraturas vertebrais de T4–L4 em indivíduos não escolióticos.³²

Subsequentemente, surgiram variadíssimos estudos que avaliaram o potencial desta ferramenta

no diagnóstico de fraturas de outras regiões anatómicas. Numa perspectiva de medicina preventiva, a IA tem sido uma mais-valia. Modelos de ML têm sido utilizados na predição de risco de quedas e fraturas com algoritmos como *CatBoost*, *SVMs* e *XGBoost*, que demonstraram superioridade relativamente a ferramentas convencionais, tais como o FRAX, na previsão de fraturas por fragilidade.^{45,46}

Já no período perioperatório, Kamer et al. aplicaram modulação 3D com base em tomografias computadorizadas para avaliar o corredor ósseo da articulação sacroilíaca e planejar o posicionamento de implantes. Esta tecnologia é aplicada atualmente em vários procedimentos, incluindo fixação de ossos longos, substituição articular, correção de deformidades da coluna e correções pós-traumáticas.⁴⁷

Com a utilização do planeamento cirúrgico assistido por IA, observou-se uma redução no tempo de aprendizagem dos cirurgiões, comparativamente a métodos tradicionais. Adicionalmente verificou-se também num aumento do sucesso das intervenções que adviu desta utilização. Paralelamente, no campo da regeneração óssea e engenharia tecidual, abordagens baseadas em ML têm sido utilizadas para modelar e prever padrões de regeneração celular.⁴⁸

3. Inteligência Artificial na deteção de fraturas

3.1. Epidemiologia e Aplicabilidade de modelos de Inteligência Artificial

O panorama epidemiológico das fraturas ósseas revela uma acentuada progressão tanto no número absoluto de casos quanto na carga de incapacidade inerente. Verificou-se que em 2019, a nível global, foram registados 178 milhões novos casos de fraturas. Este número representa um aumento significativo de 33,4% desde 1990.¹⁴ Este incremento poderá em parte ser reflexo de crescimento populacional aliado a um progressivo envelhecimento da mesma. No que tange à distribuição por género, denota-se uma maior propensão a fraturas no sexo masculino em fases mais jovens da vida, com uma inversão da tendência em faixas etárias mais velhas, havendo uma incidência superior de fraturas em mulheres num período pós-menopausa. Tal disparidade estatística pode ser explicada por fatores de índole ocupacional, bem como a padrões comportamentais.^{14,49}

Posto isto, e tendo em conta o supramencionado acerca da sobrecarga e subdiagnóstico, diversos estudos têm explorado diferentes metodologias para aferir desempenhos e identificar espaços de melhoria na forma como a IA é usada na deteção de fraturas.⁵⁰ É, contudo, de suma importância compreender entre eles, quais modelos têm demonstrado melhores desempenhos e *nuances* diferenciadores e vanguardistas. Dentro das métricas de performance do mesmo, é tanto

de valorizar as mesmas em termos absolutos quanto relativas ao desempenho humano e explorar o potencial de uma sinergia entre ambos.

Dada a extensa literatura e ausência de consensualidade entre si sobre o tema, de forma a perceber que regiões anatómicas seriam mais pertinentes avaliar, combinaram-se dados epidemiológicos de incidência e anos vividos com incapacidade de diferentes tipos de fraturas e concluiu-se que as fraturas apendiculares são aquelas que se apresentam como mais incidentes e com maior impacto anos vividos com incapacidade. Este dado, em combinação com a disponibilidade de estudos de IA nesta área, fez com que as áreas anatómicas de maior enfoque na presente revisão, fossem a região da anca e do punho.¹⁴

É de ressaltar que, apesar da seleção destas regiões anatómicas em virtude da sua vasta validação e presença recorrente na literatura, o desempenho da IA como ferramenta de diagnóstico encontra-se igualmente demonstrado noutras áreas do corpo humano. Já tendo sido alvo de estudo regiões como a mandíbula, coluna vertebral, costelas, cotovelo, joelho e articulação tibiotársica, onde os modelos evidenciaram desempenhos diagnósticos comparáveis e, por vezes, superiores aos de clínicos.⁵¹⁻⁵⁸

Na região mandibular, Warin et al. aplicaram modelos como o *YOLOv5* e o *Faster R-CNN*, obtendo métricas de *Area Under the Curve* (AUC) próximas de 0,90, com desempenhos superiores aos de cirurgiões maxilofaciais.⁵¹ Na coluna vertebral, Murata et al. utilizaram o modelo *Visual Recognition V3*, alcançando uma acurácia de 86,0%, sensibilidade de 84,7% e especificidade de 87,3%, com desempenhos não inferiores aos de cirurgiões ortopédicos.⁵² Já Inagaki et al. testaram diversas arquiteturas, como *InceptionV3*, *Xception* e *ResNet*, atingindo, no caso do *InceptionV3*, uma precisão média de 98,9% e uma sensibilidade de 86,0%, demonstrando desempenho superior ao de especialistas na deteção de fraturas sacrais.⁵³

No tórax, Wu et al. recorreram ao *YOLOv3* para identificar fraturas costais, obtendo uma sensibilidade de 91,1% e uma precisão média de 81,6%, evidenciando o potencial da IA como ferramenta de triagem.⁵⁴ Na região do cotovelo, Choi et al. relataram diagnósticos precisos de fraturas supracondilíneas pediátricas com desempenho equivalente ao de radiologistas.⁵⁵ No joelho, Liu et al. aplicaram o modelo *RetinaNet* para detetar fraturas dos pratos tibiais, alcançando uma precisão de 91,0% e uma velocidade de análise cerca de 16 vezes superior à dos médicos.⁵⁶ Por fim, na região do tornozelo, Ashkani-Esfahani et al. demonstraram que o *InceptionV3* atingiu sensibilidades e especificidades superiores a 98,0% na deteção de fraturas, enquanto Prijs et al. verificaram que modelos como o *Mask R-CNN* conseguem discriminar eficazmente fraturas em casos simples e complexos, com especificidades próximas de 96,0% e acurácias a rondar os 90,0%.^{57,58}

3.2. Inteligência Artificial na detecção de fraturas do punho

3.2.1. Epidemiologia das fraturas do punho

O punho, a par com a mão, constituem uma das regiões anatómicas mais frequentemente fraturada, com uma incidência cumulativa global superior a 18 milhões de fraturas anuais a nível global.^{59,60} As fraturas do rádio distal (FRD) e as do cúbito são as mais prevalentes.⁶¹ Considerando o carpo, o escafoide, é aquele que se encontra mais frequentemente fraturado (89% das fraturas do carpo), com incidência anual estimada de 5 fraturas por 10.000 pessoas.⁶² As FRD representam cerca de 20% de todas as fraturas, condicionando a utilização não menosprezável de recursos em contexto de serviço de urgência.⁶¹ Denota-se uma prevalência superior em indivíduos de meia-idade e idosos com osteoporose, resultando frequentemente de quedas (traumatismos de baixa energia), por oposição em indivíduos mais jovens onde surgem mais associadas a traumatismos de alta energia.⁶³ Uma vez que representam uma causa de ida frequente ao SU, acabam muitas vezes a ser diagnosticadas por médicos não especialistas em ortopedia ou radiologia, a quem cabe tomar decisões clínicas imediatas. Dadas as circunstâncias, não é assim tão infrequente haver erros de diagnóstico, nomeadamente no não reconhecimento de fraturas.⁵⁰

A ausência de diagnóstico atempado pode levar a atrasos no tratamento, consolidações viciosas e desenvolvimento de osteoartrose, o que por sua vez se repercute em diminuição da recuperação funcional.^{64,65} Este risco é particularmente elevado em indivíduos idosos, cuja recuperação é geralmente mais lenta e sujeita a complicações devido à menor reserva funcional e capacidade de regeneração dos tecidos.⁶⁵

3.2.2. Performance de modelos disponíveis

Diversos estudos de avaliação de *performance* para esta região anatómica comparam os diferentes algoritmos disponíveis (tabela I). As diferenças não residem apenas nas escolhas de arquiteturas computacionais (como *U-Net*, *EfficientNet*, *RegNet*, *InceptionResNet*), mas também nas abordagens de preparação da imagem, estratégias de segmentação, nível de supervisão dos dados e, finalmente, no tipo de diagnóstico desejado. Este último abarca um leque variado de resultados, que se estende desde a mera deteção binária da presença ou ausência de fratura até à sua caracterização detalhada segundo sistemas de classificação validados, próprios ao tipo específico de lesão em análise. Para além disso, torna-se possível recorrer a técnicas como o *Grad-CAM*, que permitem identificar as regiões da imagem que mais influenciam a decisão do modelo, ou então empregar modelos baseados na deteção de objetos, com o intuito de localizar a área anatómica afetada mediante o recurso a *bounding boxes* ou à segmentação precisa das regiões fraturadas.⁶⁶⁻⁶⁸

No que concerne ao pré-processamento de imagens e à forma como este impacta diretamente os resultados dos modelos IA, Erdaş et al. propõem uma abordagem bifásica para a detecção de fraturas do cúbito e rádio. A metodologia em causa contemplou a utilização de três modalidades de imagens radiográficas: originais, sem qualquer edição; submetidas à remoção do fundo radiográfico; e sujeitas à eliminação seletiva dos tecidos moles, conservando exclusivamente as estruturas ósseas para análise algorítmica. Testou três modelos de DL com os diferentes tipos de preparação de imagem, *RegNetX006*, *EfficientNet B0* e *InceptionResNetV2*, que procederam então à detecção de presença ou ausência de fraturas. Dentro destes, o modelo *RegNetX006*, combinado com o pré-processamento com remoção tecido não ósseo, evidenciou um desempenho substancialmente superior aos demais, alcançando acurácia de 99,2%, sensibilidade de 99,7% e especificidade de 98,6%. Este resultado permitiu não só demonstrar a eficácia de *RegNetX006*, mas também a importância do processamento de imagens, dado que se verificou uma melhoria proporcional do desempenho à medida que se aumentava o grau de pré-processamento.⁶⁹

Ainda no domínio pré processamento de imagens, Anttila et al. desenvolveu um modelo de segmentação de FRD desenhado com base na arquitetura *U-Net*, já usada previamente em domínios médicos com este mesmo intuito. A segmentação consiste num processo que procura delimitar e diferenciar regiões anatómicas ou funcionalmente distintas dentro de uma imagem.⁷⁰ Seguiu uma abordagem *patch-based*, na qual cada imagem foi dividida em subimagens de 130×130 pixels, assegurando uma sobreposição correta nas extremidades. Desta forma, pôde analisar regiões localizadas da imagem, o que poderá em parte melhorar a detecção de estruturas relevantes. As predições obtidas por cada subimagem foram posteriormente combinadas para formar um mapa de confiança de fratura referente à imagem radiográfica integral. Para o conjunto global de radiografias, a rede obteve uma acurácia de 91,0%, uma sensibilidade de 92,0% e uma especificidade de 88,0%. Da combinação de projeção PA e lateral de cada utente, resultou uma acurácia de 94,0%, sensibilidade de 95,0% e especificidade de 91,0%. Apesar das métricas estatísticas se revelarem inferiores ao estudo supramencionado, destaca-se a capacidade de segmentação por parte deste modelo, com uma detecção concreta das zonas de fratura e não baseada apenas na classificação binária.⁶⁷

Min et al., por sua vez, introduz uma variante na apresentação dos resultados, com a classificação estrutural de FDR em intra-articulares ou extra-articulares. Utilizando também uma abordagem em duas etapas, os autores integraram um *ensemble* de modelos, com *YOLOv5* para a detecção *Region of Interest* (ROI), seguido de um classificador baseado em *EfficientNet-B3*. O modelo *YOLOv5* foi avaliado através da *intersection over union* (IoU) entre a ROI predita e a *ground truth* (Fig.3), e uma vez combinado no modelo *ensemble*, demonstrou atingir um IoU médio de 81,6% na detecção da ROI. As fraturas localizadas nas ROI foram subsequentemente classificadas como intra-

articulares ou extra-articulares através de um *ensemble* de redes *EfficientNet-B3*. A agregação das predições dos dez modelos base resultou numa AUC de 0,81, com acurácia global de 81,0% e sensibilidade de 83,0%. É de destacar ainda que, em média, a análise de cada imagem demorou apenas 2 segundos. A análise visual assistida por Grad-CAM evidenciou a ativação nas zonas fraturadas, nomeadamente na interface entre a linha de fratura e a superfície articular, contribuindo para a interpretabilidade dos resultados.⁶⁶

Por fim, Yoon et al. oferece um contributo especialmente relevante ao abordar a deteção de fraturas do escafoide, com especial ênfase nas fraturas ocultas, que constituem um grande desafio diagnóstico na prática clínica.⁷¹ Com este objetivo, à semelhança de Min et al., foi desenvolvido um modelo de duas etapas, onde num primeiro momento, as imagens radiográficas foram recortadas na região do escafoide, sendo seguidamente aplicadas duas CNNs baseadas em *EfficientNetB3*, treinadas separadamente para a identificação de fraturas visíveis e ocultas. Combinou então uma arquitectura *Cascade R-CNN*, treinada de forma a isolar automaticamente a região anatómica do escafoide em radiografias da mão, com *EfficientNetB3* na classificação. A arquitetura *EfficientNetB3* foi selecionada em detrimento de outras, com base em evidências de desempenho superior face a modelos anteriores como o *Inception v3*.⁷² A classificação foi baseada nas conclusões de dois modelos distintos, o primeiro, treinado apenas com imagens de fraturas aparentes; e o segundo, treinado apenas com imagens de fraturas ocultas confirmadas através de RM. As saídas do modelo incluíam tanto a previsão da presença de fratura, como também mapas *Grad-CAM*, que permitiam interpretar visualmente as regiões com maior probabilidade de lesão. No que tange aos resultados, o modelo para fraturas visíveis apresentou desempenho com AUC de 0,96, sensibilidade de 87,1% e especificidade de 92,1%. A segunda rede, dedicada às fraturas ocultas, alcançou uma AUC de 0,81 e conseguiu identificar corretamente 90,9% das lesões invisíveis a olho nu, previamente diagnosticadas apenas através de outras modalidades imagiológicas.⁴⁰

3.2.3 Performance de Inteligência Artificial vs. Médico

Numa perspetiva diferente, que não visa apenas a avaliação da performance de diferentes modelos e variantes que influenciam os seus resultados, mas o potencial destes numa possível combinação com clínicos, serão seguidamente alvos de análise estudos que se focam nesta perspetiva combinada na avaliação de fraturas do punho.

Neste sentido, Lee et al. desenvolveu um modelo com capacidade de deteção de três tipos de fratura do punho: FRD, da apófise estilóide cubital e escafoide. O modelo foi testado de forma autónoma, mas também como ferramenta de apoio a dois radiologistas ainda em formação. Combinou um *ensemble* de 3 modelos (*RetinaNet*, *DeepLabv3*, *NasNet*) com o objetivo de identificação das ROI (rádio, apófise estiloide do cúbito e escafoide), segmentação e classificação

da fratura, respetivamente. A decisão final foi estabelecida mediante a comparação entre os valores probabilísticos oriundos da análise de segmentação e da análise de classificação, ambos confrontados com seus respetivos limiares pré-determinados para a presença de fratura. O modelo obteve resultados de AUC de 0,90 para o rádio distal, 0,93 para a apófise estilóide do cúbito e 0,81 para o escafoide. Aquando da utilização combinada por parte de radiologistas menos experientes, a AUC para deteção de fraturas foi especialmente incrementada na região do escafoide, comparativamente à sua performance individual sem o auxílio da ferramenta de IA. Verificou-se um aumento na AUC de 0,75 para 0,85 no primeiro radiologista e de 0,71 para 0,80 no segundo radiologista. Esta combinação, culminou no aumento do tempo de leitura em 110% ($p=0.007$), todavia, o acréscimo absoluto no tempo foi inferior a um segundo. Assim, apesar de estatisticamente significativo, não tem propriamente impacto na prática clínica. Não obstante, em estudos anteriores, como em Hendrix et al., não se verificou este aumento de tempo de leitura, pese embora neste estudo mencionado tratar-se de profissionais com vários anos de experiência.^{43,71}

Gan et al. também propôs uma solução baseada na integração de dois modelos distintos: o *Faster R-CNN* para deteção automática das ROI e o *Inception-v4* para a classificação das FRD. O desempenho combinado das arquiteturas mostrou que o *Faster R-CNN* identificou corretamente todas as ROI, enquanto o *Inception-v4* demonstrou acurácia de 93,0% e AUC de 0,96 na distinção entre punhos fraturados e normais. Comparativamente, o modelo superou o desempenho de radiologistas e equiparou-se ao de ortopedistas. A principal limitação metodológica foi o uso exclusivo de radiografias antero-posteriores (AP), o que pode comprometer a generalização para contextos clínicos onde imagens laterais são essenciais para avaliação tridimensional da fratura.⁷³

Suzuki et al, por sua vez, abordou esta lacuna ao treinar modelos de IA com imagens de projeções AP e lateral, com o objetivo de detetar FRD. Durante o processo de treino, o algoritmo *EfficientNet-B2* na projeção AP e a *EfficientNet-B4* na lateral, foram aqueles que entre um conjunto de modelos, mostraram maior acurácia no set de validação, sendo por isso posteriormente usados nas imagens combinadas das duas projeções. O melhor desempenho final por parte da IA registou-se para o *ensemble* no uso de projeções combinadas, com acurácia de 99,3%, sensibilidade de 98,7% e especificidade de 100%, e AUC de 0,99. Estes resultados ultrapassaram os de três ortopedistas que participaram no estudo. Contudo, o modelo opera apenas como classificador binário e não localiza fraturas, o que pode limitar a sua aplicabilidade prática. Ademais, o estudo restringiu-se a radiografias no adulto, não avaliando a performance em exames pediátricos, onde a presença de cartilagens de crescimento pode dificultar o diagnóstico.⁷⁴ Comparando este estudo com Blüthgen et al. que seguia a mesma base metodológica, percebe-se que este atingiu métricas bastante inferiores, a registar uma sensibilidade de 78,0%, especificidade de 82,0% e uma área sob

a curva (AUC) de 0,89. Tendo sido neste último, os parâmetros de performance comparáveis aos obtidos por radiologistas.⁷⁵

Zhang et al. em muito se assemelhou aquando da utilização de um modelo *ensemble* composto por três arquiteturas (*Faster R-CNN*, *RetinaNet*, *Cascade R-CNN*) de DL para detetar FRD em radiografias AP e laterais. Na utilização de imagens AP e laterais em conjunto do mesmo doente, obteve valores de acurácia, sensibilidade e especificidade de 97,6%, 98,4% e 96,7%, respetivamente, superando significativamente os resultados dos médicos assistentes de ortopedia e radiologia.⁷⁶

Tal como Yoon et al., que explorou o desempenho de IA em fraturas do escafóide visíveis e ocultas, também Langerhuizen et al se debruçou neste sentido, comparando, adicionalmente, o desempenho da IA com o de cinco ortopedistas. *VGG16*, treinado com apenas 300 casos, alcançou uma AUC de 0,77, acurácia de 72,0%, sensibilidade de 84,0% e especificidade de 60,0%. Ainda que inferior aos observadores humanos em termos de especificidade e acurácia, com os cirurgiões a registarem valores de 93,0% e 84,0%, destaca-se que o algoritmo detetou cinco das seis fraturas ocultas presentes no conjunto de teste, um resultado superior ao dos clínicos, que detetaram apenas três entre si. O desempenho relativamente inferior da IA pode ser atribuído ao pequeno tamanho da amostra. O estudo conclui que, embora a IA ainda não substitua a análise humana, possui potencial como ferramenta de triagem para suspeitas clínicas, especialmente quando se pretende decidir sobre a necessidade de exames complementares como TC ou RM.⁴⁰

3.3. Inteligência Artificial na deteção de fraturas da anca

3.3.1. Epidemiologia das fraturas da anca

As fraturas da anca revestem-se de especial importância, dada a sua prevalência na população geriátrica, que se traduz numa das principais causas de morbilidade e mortalidade a nível mundial. Nos Estados Unidos da América, registaram-se aproximadamente 250.000 episódios de fratura desta região, por ano.⁷⁷ Estão frequentemente associadas a uma acentuada deterioração funcional, bem como a uma taxa de mortalidade considerável, podendo esta atingir os 30% no decurso do primeiro ano após a ocorrência.⁷⁸ Não obstante, a acuidade diagnóstica apresenta limitações significativas, com taxas de erro estimadas entre os 4% e os 9%, o que pode influenciar fortemente o prognóstico, incluindo progressão para doença articular degenerativa, osteoartrite secundária e perdas irreversíveis de mobilidade. O descrito, traduz-se inevitavelmente, numa

deterioração marcada da qualidade de vida e um acréscimo substancial dos encargos económicos para os sistemas de saúde.⁵⁰

Importa salientar ainda que dado o envelhecimento populacional, justifica-se o incremento no número de artroplastias totais da anca que contabilizam um total de mais 90.0000 procedimentos efetuados anualmente, no Reino Unido.⁷⁹ Como resultado, verifica-se também um crescimento nas complicações pós-operatórias, destacando-se entre elas as fraturas periprotésicas do fémur (FPF), que afetam cerca de 3,5% dos pacientes submetidos a esse tipo de cirurgia.⁸⁰ Ocorrem principalmente devido a quedas de baixa energia em pacientes idosos, mas também podem ser causadas por descelagem da prótese, osteólise ou sobrecarga mecânica proveniente de um implante adjacente.⁸¹ Para orientar a conduta cirúrgica, utiliza-se frequentemente o *Vancouver Classification System* (VCS), que classifica as fraturas com base em três critérios principais: localização da fratura, estabilidade do implante e qualidade óssea. As fraturas Tipo A envolvem o grande ou pequeno trocânter, as Tipo B estão localizadas adjacentes ao implante, enquanto as Tipo C ocorrem abaixo da haste femoral.⁸²

3.3.2. Performance de modelos disponíveis

Neste registo de avaliação de performance, tal como efetuado em 3.2.2, proceder-se á à análise do que existe para a região da anca (Tabela II).

Neste contexto, surge na vanguarda Kim et al., que, de forma progressista comparativamente a trabalhos anteriores como Zdosek et al., não se limitou apenas à criação de um algoritmo que detetasse fraturas atípicas completas do fémur (FAF), como também procurou viabilizar o diagnóstico precoce dessas fraturas enquanto ainda se encontram em fase incompleta.⁸³ As FAF têm maior incidência em doentes com curvaturas excessivas do fémur ou que têm terapêutica instituída com bisfosfonatos. Decorrem de um processo de microfratura, que, após uma tentativa de cicatrização, volta a ocorrer repetidamente, sendo que a perpetuação deste ciclo contínuo de microfraturas e cicatrização provoca deformações no córtex lateral do fémur, resultando na fragilidade do osso.⁸⁴ Embora possam ser identificadas através de cintigrafia óssea ou na apresentação de edema na medula óssea na RM, apresentam um alto índice de erro diagnóstico, o que pode levar a atrasos no diagnóstico, permitindo que uma fratura incompleta evolua para uma forma completa.⁸⁵

Após a aplicação do filtro Sobel, empregue no sentido de realçar as bordas ósseas, foi realizado o treinamento de seis modelos amplamente reconhecidos na literatura, utilizando a técnica de *transfer learning*. Os modelos selecionados foram: *EfficientNet B5, B6, B7; DenseNet121; e MobileNet V1 e V2*. A arquitetura do sistema foi organizada em duas abordagens de *ensemble*, sendo uma composta pelos três modelos de maior acurácia e outra utilizando os cinco melhores. A

abordagem *Ensemble Top 3* demonstrou desempenho ligeiramente superior, com uma AUC de 0,998. Várias conclusões frutíferas advêm deste estudo, como o cunho na importância do processamento de imagens, com a aplicação do filtro de Sobel a incrementar até 5% na acurácia dos modelos avaliados. Verificou-se ainda que a utilização de um *ensemble* restrito a modelos de elevada precisão proporcionou desempenho superior em comparação com combinações que incluíam modelos de menor acurácia. Entre os modelos testados, o *MobileNet V1* destacou-se como o mais eficiente em termos de desempenho computacional, sendo capaz de processar 28 radiografias por segundo. Por outro lado, o *ensemble* constituído pelos três melhores modelos apresentou uma taxa de processamento inferior, com cerca de 5 radiografias analisadas por segundo, apesar de ter proporcionado uma melhoria de aproximadamente 1% na acurácia em relação a modelos isolados. Uma vez testado num conjunto externo, o modelo ensemble alcançou uma acurácia de 92,5%, que incluiu imagens provenientes de várias fontes e com variabilidades significativas e com imagens de qualidade aquém das utilizadas na prática clínica real, o que atesta a modelo face à heterogeneidade dos dados.⁸⁶

No sentido de perceber o papel que a IA poderá ocupar na deteção de FPF, Alzaid et al. analisaram uma *coorte* de doentes submetidos a artroplastia total da anca. Avaliou tanto a classificação binária (presença ou ausência de fratura) como a classificação multi-classe (tipos de fraturas conforme o VCS), bem como a classificação e localização das fraturas por algoritmos especializados na deteção de objetos. Desta forma numa primeira fase, testou-se a classificação binária, que consistia na diferenciação entre radiografias com e sem fratura, tendo para isso testado quatro modelos CNN: *Inception*, *VGG*, *ResNet50* e *DenseNet161*. Os modelos *Densenet161*, *Resnet50* e *Inception* apresentaram acurácia semelhante, de cerca de 95% na classificação binária. Sendo atualmente dos poucos estudos que se propõem à pesquisa deste tipo de fratura, e tendo por exemplo, como termo de comparação Miao et al, que apenas se focou na deteção de fraturas no fémur sem presença de próteses, estes resultados revelam-se promissores na medida em que ultrapassam a acurácia deste de apenas 91,0%.^{87,88} Todavia, o mesmo não se verificou aquando da classificação multiclasse. No cômputo geral, o *Resnet50* foi o que apresentou o melhor desempenho na classificação dos tipos de FPF, com uma acurácia média de 90%. Apresentou, contudo, valores baixos de sensibilidade na classificação do Tipo A, o que pode derivar do menor número de imagens deste tipo. Ademais, na distinção entre os Tipos B e C também obteve um desempenho ligeiramente inferior, uma vez que estes dois tipos de padrões são semelhantes entre si, por vezes. Já nos modelos de localização de objetos, *Faster RCNN* e *RetinaNet*, verificaram-se métricas inferiores, com melhor desempenho do primeiro, que mesmo assim ficou aquém dos modelos classificadores, com acurácia de 78,0% mas com sensibilidade de 98,0%.⁸⁷ Numa tentativa de perceber o impacto do pré processamento com o fornecimento de imagens completas como

entrada *versus* imagens demarcadas nas zonas de interesse, obtiveram resultados interessantes e dispares dos demais, uma vez que a *performance* do modelo em muito se assemelhou nas duas modalidades de fornecimento de imagem.⁸⁷

Ouyang et al. distingue-se dos demais estudos que visaram o desenvolvimento de um modelo de CNN na detecção de fraturas do fêmur em radiografias pélvicas, por estruturar a criação do mesmo com base em princípios do *design thinking* — Empatia, Definição, Ideação, Prototipagem e Teste — assegurando uma correspondência efetiva entre ferramenta proposta e as necessidades reais da prática clínica. Numa primeira fase foram efetuadas entrevistas junto de ortopedistas especializados em trauma, que evidenciaram a problemática da elevada taxa de erros na identificação de fraturas em imagens pélvicas. A partir daí sucedeu-se a formulação do problema clínico, orientando o desenvolvimento subsequente do modelo. Durante a fase de Ideação, explorou-se a integração da IA no fluxo clínico, culminando na conceção de um modelo baseado na arquitetura *Xception*. Para promover a interpretabilidade do sistema, incorporou-se a técnica *Grad-CAM*, possibilitando a visualização das regiões da imagem mais determinantes para as decisões do algoritmo. Fomentou-se ainda a interação direta entre médicos e o sistema, recolhendo feedback qualitativo por parte destes no sentido de melhorar a ferramenta, o que culminou em última instância na substituição da classificação binária por um mapa visual interativo das áreas suspeitas de fratura. Traduziu-se numa elevação da acurácia de 91,0% para 95,0% e a especificidade de 84,0% para 93,0%, após otimização do modelo tendo em conta o feedback dos clínicos.⁸⁹

3.3.3. Performance de Inteligência Artificial vs. Médico

Novamente numa vertente comparativa, de forma a perceber o potencial que este tipo de ferramentas pode apresentar no papel auxiliar a clínicos, analisar-se-á seguidamente os estudos que se focaram sobre a sinergia entre a Inteligência artificial e a humana.

Urakawa et al. testa a eficácia da arquitetura *VGG-16* na detecção de fraturas intertrocantéricas. Desta forma, o foco metodológico incidiu na análise de radiografias recortadas manualmente para isolar o fêmur proximal, mitigando interferências anatómicas. De uma forma individual o modelo atingiu uma acurácia global de 95,5%, que se verificou ser superior à dos ortopedistas que se fixou nos 92,2%. A título de performance individual dos modelos, destaca-se de estudos anteriores, como Olczak et al., que apesar de se sucumbir da mesma arquitetura obteve uma acurácia inferior, de 83,0% na identificação de fraturas em radiografias da mão, punho e tornozelo e em muito comparável ao dos radiologistas humanos (82,0%).⁴² A superioridade evidenciada em Urakawa poderá ser, em parte, ser atribuída à estratégia metodológica que privilegiou a análise de imagens recortadas do fêmur proximal, eliminando potenciais interferências

provocadas por estruturas anatómicas adjacentes, como o osso pélvico ou a diáfise femoral, que poderiam comprometer a precisão do processo de classificação automatizada.⁹⁰

No sentido de comparar performances, mas com *nuances* metodológicas relevantes, Cheng et al. emprega o algoritmo *PelviXNet*, cuja arquitetura se fundamenta em redes *DenseNet-169* e *Feature Pyramid Networks* (FPN) e adota uma estratégia de anotação inovadora, privilegiando a marcação por pontos em detrimento das convencionais *bounding boxes*. Tal escolha metodológica demonstra particular pertinência face à multiplicidade e complexidade anatómica das patologias investigadas em radiografias pélvicas. Ademais, corroborou novamente a importância de um fornecimento de imagens de treino anotadas para o desempenho do modelo, uma vez que números diferentes de imagens anotadas, se traduziram em valores diferentes de performance, com AUC a subir de 0.93 para 0.97 para 1040 e 5204 imagens anotadas, respetivamente. Obteve uma acurácia global para todas as lesões de 92,4%. Sublinha-se ainda que metodologicamente se diferencia, uma vez que a maioria dos estudos treina os modelos com imagens selecionadas artificialmente a partir de bases de dados e, por isso, inevitavelmente concorrem a um risco de enviesamento. Neste caso, tem a particularidade de usar dados com o padrão de distribuição de uma população em tempo real nas idas ao Serviço de Urgência durante um ano. Adicionalmente, comparou *PelviXNet* com 22 médicos apenas na deteção de fraturas da anca e da pélvis tendo-se verificado acurácias de 99,5% e 94,5%, respetivamente. Apresentou um desempenho superior ao de médicos do serviço de urgência e internos e desempenho comparável a especialistas em Ortopedia e Radiologia.⁹¹

Kroque et al. combina *DenseNet-169* para a tarefa de classificação e uma arquitetura *ResNet-50* acoplada as FPN para a deteção de ROIs. A taxonomia de classificação abrangeu seis categorias clínicas: normalidade, fratura do colo femoral deslocada e não deslocada, fratura intertrocantérica, redução aberta e fixação interna prévia ou artroplastia. No que concerne ao desempenho diagnóstico num teste de validação interna, o modelo evidenciou uma acurácia binária de 93,7% empregando marcações manuais, registando um incremento marginal para 94,2% aquando da integração com um modelo de geração automática de *bounding boxes* nas ROIs.

A comparação direta com especialistas de radiologia e ortopedia subespecializados na região da anca, utilizou um conjunto de subteste de 100 imagens. O desempenho foi estatisticamente superior no modelo, particularmente na comparação da análise de imagens de baixa resolução, cuja utilização por parte do modelo se impôs por limitações computacionais. Nestas condições de menor qualidade imagiológica, os especialistas alcançaram 88,6% de acurácia binária e 83,4% em classificação multiclasse, ao passo que o modelo de IA atingiu 95,8% e 92,8%, respetivamente. A análise de imagens de alta resolução permitiu uma aproximação do desempenho dos especialistas aos valores obtidos pela IA (93,5% em classificação binária e 89,9% em multiclasse). Em contrapartida, os internos demonstraram um desempenho substancialmente

inferior em ambas as resoluções imagiológicas. Verificou-se ainda um incremento no desempenho diagnóstico observado tanto em especialistas quanto em internos aquando da colaboração com a IA. Adicionalmente, a robustez do *ground truth*, alicerçada não apenas em radiografias simples, mas também em exames imagiológicos como em TC, RM e imagens pós-operatórias, mitiga as limitações impostas pelo reduzido tamanho amostral e confere elevada fiabilidade aos rótulos atribuídos, prevenindo potenciais enviesamentos decorrentes da ocultação de fraturas em radiografias convencionais, que ronda os 10% no caso de fraturas da anca como descrito em Dominguez et al.^{92,93}

A relevância de uma *ground truth* comprovada por outras modalidades imagiológicas também foi evidenciada em Tsubasa Mawatari et al., que da mesma forma utilizou a TC e RM como referência. Esta abordagem assegura uma rotulagem de dados mais precisa e fiável, garantindo um critério diagnóstico mais rigoroso. No entanto, este método pode comprometer os valores de desempenho do modelo, contrastando com a grande maioria das investigações precedentes que se guiam em relatórios radiológicos ou padrões de referência estabelecidos retrospectivamente por especialistas, com base exclusiva nas radiografias, o que pode resultar numa inclusão de casos falsos negativos desde o início do estudo. Através do *Inception V1* atingiu uma AUC de 0,905, equiparável ao desempenho dos leitores especialistas (sendo a AUC do Leitor 2 com 22 anos de experiência de 0,920).⁹⁴ Não sendo possível uma inferir uma conclusão direta, dadas as diferenças metodológicas, constata-se ser um valor inferior a estudos anteriores que utilizaram imagens radiológicas apenas como *ground truth*.^{35,90,95}

Na linha de trabalho em classificação multiclasse, Jiménez-Sánchez et al., reconhecendo a intrínseca dificuldade e a limitada concordância interobservador na classificação AO (Fig.4) dos padrões de fratura do fémur proximal, investigou três metodologias distintas de classificação automática destas: análise da imagem radiográfica completa; classificação baseada em ROIs definidas manualmente por especialistas; classificação com localização automática da fratura. A classificação foi efetuada por *ResNet-50* pré-treinada no *ImageNet*, enquanto a localização da fratura foi efetuada por um modelo *AlexNet*, demonstrando elevada precisão na previsão dos centros das ROIs. O estudo evidenciou que a utilização da imagem radiográfica completa proporcionou uma acurácia média de 89,0% na classificação multiclasse, com sensibilidade a oscilar entre 67,0% para fraturas tipo A e 95,0% para casos normais. A introdução de ROIs manuais elevou a acurácia para 91%, com particular incremento na sensibilidade para fraturas tipo A (75,0%), o que está de acordo com outros estudos neste sentido.⁹² A abordagem de localização automática, por sua vez, alcançou resultados equiparáveis, com acurácia binária de 93,0% e multiclasse de 90,0%. Comparativamente, o desempenho de três médicos (especialistas em Ortopedista, Radiologia e Interno de especialidade) que obtiveram uma acurácia média de 91% e sensibilidade de 87,0%,

embora com um tempo médio de análise de 46 minutos, contrastando com a execução em tempo real da ferramenta de IA.⁹⁶

Por fim, destaca-se ainda pela abordagem multidisciplinar, uma vez que não se foca apenas no desenvolvimento de um sistema de diagnóstico assistido por computador para a análise automatizada de radiografias do fêmur proximal, mas também no potencial de aplicabilidade de ferramenta de IA como contributo pedagógico para ortopedistas em formação. Em conformidade com isso foi desenvolvido um módulo de recuperação de imagens semelhantes, com vista à aplicação pedagógica. Este mecanismo baseia-se na extração de vetores de características da penúltima camada da rede neural e na medição de distâncias euclidianas, permitindo a recuperação de casos análogos. Em testes realizados, 70% das imagens recuperadas pertenciam à mesma classe da imagem de referência.⁹⁶

Também na abordagem multiclasse, Leonardo Tanzi et al. propôs uma abordagem hierárquica com base na classificação AO/OTA (Fig. 5) das fraturas do fêmur proximal, recorrendo à arquitetura *InceptionV3*. À semelhança de Jiménez-Sánchez et al., o estudo excluiu as fraturas do tipo C devido à escassez de exemplos dessa categoria, concentrando-se unicamente nos tipos A e B. A principal distinção metodológica residiu na subclassificação detalhada do tipo A, segundo a tipologia AO/OTA: A1 (pertrocanterica simples), A2 (pertrocanterica multifragmentar com parede lateral incompetente) e A3 (intertrocanterica). Para refletir o processo de raciocínio clínico inerente à classificação AO/OTA, os autores adotaram um modelo multietapas. Na tarefa de triagem entre fraturas dos tipos A, B e casos não fraturados, o modelo *InceptionV3* atingiu uma acurácia de 87,0%. Quando submetido à classificação mais refinada entre A1, A2, A3, B e normal, obteve 78,0% de acurácia. Ao aplicar uma rede em cascata numa segunda abordagem, registou-se uma melhoria de desempenho para 80,0% na classificação das cinco categorias, correspondendo a um acréscimo de 2% face à abordagem inicial. A utilidade do modelo enquanto ferramenta de apoio à prática radiológica também foi avaliada, revelando-se vantajosa, sendo que se observou-se uma elevação de 14% na acurácia diagnóstica e uma redução significativa no tempo de análise. Contudo, apesar do modelo *multistage* procurar replicar a lógica diagnóstica do clínico, o seu desempenho revelou-se inferior ao de Jiménez-Sánchez et al., o que em parte poderá consolidar a relevância da utilização de ROIs como fator determinante para uma classificação mais precisa.⁹⁷

3.4. Inteligência Artificial na detecção de fraturas pediátricas

3.4.1. Epidemiologia das fraturas pediátricas

As fraturas pediátricas contabilizam uma apresentação bastante frequente em Serviços de Urgência. A incidência anual de fraturas na população pediátrica varia entre 9,5 a 20,2 por cada 1000 crianças.⁹⁸

Por oposição ao que se verifica no contexto da população adulta, a utilização da IA na detecção de fraturas pediátrica, não se encontra tão amplamente estudada, sendo a maioria dos estudos referentes a diferentes regiões anatómicas relativos a modelos que se encontram atualmente comercializados.^{99,100} Tal constatação poderá advir do facto da aplicação de modelos de IA em contexto pediátrico acarretar dificuldades adicionais decorrentes da complexidade morfológica do esqueleto infantil, que está sujeito a transformações contínuas ao longo do crescimento, com diversas variantes anatómicas inseridas no espectro da normalidade e padrões de fratura que diferem substancialmente da apresentação em adultos.¹⁰¹

3.4.2. Performance dos modelos de IA na detecção de fraturas pediátricas

Os estudos dedicados exclusivamente à população pediátrica revelaram resultados positivos, os quais se encontram enumerados de forma sistemática na Tabela III.

Zech et al. realizaram dois estudos consecutivos com uma metodologia semelhantes em ambos. O primeiro, publicado em 2023 concentrou-se na região do punho, utilizando o *Faster R-CNN*, adaptado por *transfer learning* a partir do conjunto *MS COCO*. O modelo foi comparado com médicos internos (Radiologia e Pediatria), individualmente e em associação com IA, destacando-se pela melhoria substancial na acurácia diagnóstica, em particular para fraturas *buckle* do rádio distal (de 69,0% para 92,0% com IA). Contudo, apresentou a grande limitação de ter uma *ground truth* estabelecida por um único médico interno com base em relatórios clínicos, e apenas se considerou a projeção AP, divergindo da prática clínica habitual.¹⁰²

No estudo subsequente (2024), Zech et al. expandiram a análise para todo o membro superior, mantendo a população pediátrica em estudo. Sem efetuar qualquer tipo de treino, e sucumbindo-se apenas de um modelo de detecção de objetos disponível, *Faster R-CNN X101-FPN*, compararam o desempenho com o de médicos internos de Radiologia e Pediatria e médico especialista em Radiologia. O modelo atingiu uma AUC de 0,89, tendo a IA melhorado substancialmente o desempenho dos internos de radiologia, sendo que AUC aumentou de 0,77 para 0,88 nos de radiologia, e de 0,71 para 0,84 nos de pediatria. Todavia, não teve impacto significativo no subgrupo dos médicos Radiologistas experientes (de 0,88 para 0,89).¹⁰³ Métricas

superiores a outros estudos, como, por exemplo, Aryasomayajula et al., que ao empregar o VGG-16 na detecção de fraturas do punho, demonstrou uma acurácia inferior (85,0%)¹⁰⁴

Ghosh et al. focaram-se na detecção de fraturas das costelas em crianças menores de dois anos. Foram avaliadas duas arquiteturas, *ResNet-50* e *DenseNet-121*, treinadas no conjunto de dados *NIH ChestX-ray14*. O desempenho foi aferido para o modelo *ResNet-50* com AUC 0,84, enquanto o *DenseNet-121* registou uma AUC de 0,82, sem diferença estatisticamente significativa entre ambos. A interpretação dos modelos foi suportada por mapas de ativação *Grad-CAM*, que confirmaram a focalização das redes nas zonas anatómicas associadas à presença de fraturas. A comparação com dois médicos especialista em Radiologia (um com cinco anos de experiência pós-residência e outro com mais de dez anos em radiologia pediátrica) evidenciou desempenho equiparável em termos de sensibilidade, embora com uma especificidade superior no radiologista mais experiente.¹⁰⁵

Binh et al., por sua vez, emprega o modelo *YOLOv4* treinado com uma base de dados que se denomina de *GRAZPEDWRI-DX* e que reúne um total de 20.327 radiografias de punho de 6.091 doentes pediátricos, vistos entre 2008 e 2018 no Departamento de Cirurgia Pediátrica da Universidade Médica de Graz, na Áustria, e que foi exatamente desenvolvida com o propósito de desenvolver a utilização de IA nesta população.¹⁰⁶ No estudo em questão, após seleção a partir dessa base, coube a um médico estabelecer a *ground truth*, desenhando uma caixa delimitadora e classificá-la de acordo com quatro classes de rótulos: *fracture of ulnar metaphysis* (FUM), *fracture of ulnar epiphysis* (FUE), *fracture of radial epiphysis* (FRE), and *fracture of radial metaphysis* (FRM). Auferiu AUCs entre 0,83 e 0,94 nas quatro categorias contabilizadas. Superou o ortopedista em determinadas categorias, como por exemplo nas fraturas da metáfise ulnar (FUM), contudo a acurácia global do algoritmo permaneceu aquém da obtida pelo radiologista experiente.¹⁰⁷

Por fim, Knight et al. introduziram uma abordagem singular ao avaliar a eficácia da ecografia 2D e 3D na detecção de FRD em crianças, bem como a aplicação de modelos de IA para interpretação automática dessas imagens. Após o exame ecográfico (realizado ainda na triagem, antes de qualquer avaliação médica), os participantes realizaram radiografias que serviram como *gold standart*, que seriam interpretadas por um radiologista pediátrico. Concomitantemente, as imagens ecográficas foram analisadas por leitores com três níveis de experiência de forma a perceber a variabilidade existente entre eles e de que forma a IA, resultando num bom desempenho, pode ajudar a colmatar tal achado.

Os modelos de IA, *ResNet34* e *DenseNet121*, foram treinados para classificar imagens ecográficas como "fratura" ou "normal". O modelo teve um bom desempenho, com *ResNet34* e *DenseNet121* a alcançarem uma melhor AUC de 0,93 e 0,91, respetivamente, para 2D e 3D. Além disso, a IA foi particularmente útil para compensar a variabilidade entre leitores

menos experientes, demonstrando desempenho comparável aos especialistas em muitos casos. A explicabilidade do modelo foi reforçada através de máscaras de segmentação coloridas, que mostravam as estruturas anatómicas e destacavam as áreas fraturadas, aumentando a confiança dos utilizadores.¹⁰⁸

4. Modelos Comerciais

Com base nas informações disponibilizadas pelo *Health AI Register* foi possível identificar um conjunto de 23 modelos comerciais de IA atualmente certificados, exclusivamente ou não, na detecção de fraturas. Para efeitos legais, alguns já obtiveram aprovação simultânea para o mercado europeu (CE) e norte-americano, enquanto outros se encontram, até à data, apenas aprovados para o espaço europeu.¹⁰⁹

Entre os modelos aplicados a radiografias, há um enfoque especial no *Gleamer BoneView®*, extensamente validado na literatura. Apresenta dupla certificação, pela CE e pela *Food and Drug Administration* (FDA). Concebido com o intuito de, não só, proceder à detecção automatizada de fraturas em diversas regiões anatómicas, como também, com a capacidade de identificar luxações e outras anomalias ósseas em imagens radiográficas. Encontra-se validado tanto em adultos como em doentes pediátricos acima dos dois anos de idade. Baseado numa CNN construída sobre a arquitetura *Detectron 2*, foi treinado com mais de 60 mil radiografias provenientes de hospitais franceses, analisando imagens de forma autónoma, com tempos de análise que vão desde 1 a 10 minutos.¹¹⁰ A sua interpretação é facilitada na medida em que demonstra diferentes níveis de confiança diagnóstica: valores superiores a 90% correspondem à classificação "positivo", delimitando a área suspeita com uma caixa preenchida; entre 50% e 90%, classifica-se como "dúvida", com delimitação tracejada; e, quando inferiores a 50%, a imagem é considerada "negativo", não sendo assinaladas regiões suspeitas. Ademais, o sistema disponibiliza uma tabela de resumo com priorização de casos urgentes, contribuindo não só para a tomada de decisão clínica, mas também para a otimização do fluxo de trabalho. Vários estudos com desenhos metodológicos distintos, desde abordagens retrospectivas, prospetivas a multicêntricas validaram a eficácia do modelo.^{110,111}

Duron et al., num estudo com mais de 600 doentes com trauma apendicular, demonstrou que a utilização deste modelo em concreto aumentou significativamente a sensibilidade e especificidade dos clínicos, ainda que a redução do tempo de leitura de 15,0% não tenha alcançado significância estatística ($p = 0,12$).¹¹⁰ Meetschen et al., por sua vez, analisando 200 radiografias de diversas regiões anatómicas, evidenciou uma melhoria substancial na sensibilidade diagnóstica

entre internos de radiologia, especialmente os menos experientes, mantendo a especificidade, e tendo neste caso atingindo uma redução do tempo com significância estatística.¹¹¹

Em relação à acuidade diagnóstica geral, acabam quase todos os estudos por convergir ao constatar que pode equiparar-se ou até superar o desempenho de radiologistas na detecção de fraturas comuns, não obstante, a sua eficácia tende a diminuir perante determinados padrões de fratura mais subtis.¹¹² Esta dicotomia está patente na avaliação da detecção de fraturas do escafoide, onde se verificou uma acurácia aceitável, contrastando com a sua limitada capacidade de identificar lesões noutros ossos do carpo.¹¹³

Avaliou-se ainda a sua aplicação no fluxo clínico, através de um estudo prospetivo em tempo real num centro de trauma. Da sua incorporação verificou-se um número considerável de correções diagnósticas efetuadas por médicos internos, sendo que após a integração dos resultados fornecidos pelo algoritmo, foram efetuadas 35 alterações diagnósticas, das quais 33 resultaram de facto em correções adequadas.¹¹⁴ Em contraste, estudos retrospectivos, como Guermazi et al., com uma prevalência de fraturas artificialmente manipulada, confirmaram uma melhoria estatisticamente significativa na sensibilidade diagnóstica em diversas regiões anatómicas, preservando a especificidade, aquando da utilização por parte de um conjunto de clínicos com especialidades diversificadas.¹¹⁵

No que concerne à população pediátrica, avaliada especificamente em estudos como Hayashi et al e Nguyen et al. apresenta uma performance bastante positiva, demonstra no último um potencial significativo, com melhorias notáveis no desempenho de médicos em formação quando coadjuvados pela ferramenta de IA.^{99,116} Não obstante, Altmann-Schneider et al. ressalva as limitações do modelo na detecção de padrões de fratura subtis e específicos, frequentemente prevalentes na prática pediátrica. A baixa sensibilidade na identificação de fraturas típicas da infância, como as fraturas em trampolim e as lesões *Salter-Harris* tipo II da tíbia proximal, bem como uma fraca detecção de fraturas de ramo verde e outras lesões específicas, como fraturas do olécrano e avulsões da apófise estilóide, evidenciam os desafios inerentes ao reconhecimento de padrões complexos e menos evidentes.¹¹⁷

Os demais modelos elencados atualmente disponíveis na comunidade europeia no panorama da IA aplicada à imagiologia médica revelam grande heterogeneidade no seu espectro funcional. Heterogeneidade essa que não se traduz apenas nas modalidades imagiológicas empregues, que vão desde a radiografia convencional, até à TC, mas também na multiplicidade anatómica das regiões avaliadas e nas diferentes das anomalias detetadas.¹¹⁶

Várias dos modelos apresentam capacidades diagnósticas que vão muito além da mera detecção de fraturas. Existem, por exemplo, modelos como o *InferRead DR Chest*® ou *Annalise Enterprise CXR*®, dedicados à análise de radiografias torácicas com aptidão para identificar não só

fraturas costais ou vertebrais, como também patologias pulmonares e pleurais, tais como derrames, atelectasias ou pneumotórax.^{118,119} Da mesma forma, sistemas especializados em TC craniana, como o qER ou o *Annalise Enterprise* CTB, são capazes de identificar fraturas cranianas associadas a hemorragias intracranianas, deslocamentos da linha média, edema cerebral entre outras emergência neurocirúrgicas.^{120,121}

Importa salientar, ademais, que alguns destes sistemas como o *TechCare Kids*® foram desenvolvidos com contexto pediátrico em mente, tendo em conta as especificidades desta faixa etária, o que minimiza o risco de enviesamento por modelos generalistas. O modelo mencionado acopla ainda funcionalidades adicionais como a estimativa da idade óssea com base em radiografias da mão e punho, com potencial apoio ao diagnóstico de patologia endócrina ou do crescimento.¹²²

O facto da deteção de fraturas representar apenas uma fração limitada no espectro de funcionalidade dos sistemas de IA mais recentes, ilustra a clara evolução desta área no campo médico, com o abandono de soluções centradas em tarefas isoladas, como a identificação de fraturas, para sistemas multifacetados que apoiam o diagnóstico clínico e são capazes de reconhecer uma vasta gama de patologias com elevada precisão, estando frequentemente adaptados para integração eficiente nos sistemas hospitalares RIS/PACS.¹⁰⁹

5. Onde se coloca a Inteligência Artificial no futuro e entraves na utilização

A integração disseminada da IA no domínio do diagnóstico de fraturas ósseas, embora represente um potencial tremendo, suscita uma miríade de considerações relevantes, que não se resumem à mera eficácia algorítmica.

Em primeiro instância, a pedra basilar do desenvolvimento de sistemas de IA reside na disponibilidade, qualidade e heterogeneidade dos dados de treino, conforme reiterado anteriormente. A performance algorítmica é intrinsecamente dependente de conjuntos de dados extensos e corretamente rotulados.¹²³ A variabilidade nas metodologias de aquisição de imagens; as disparidades institucionais nos processos de armazenamento e rotulagem de dados; a definição da *ground truth* por diferentes observadores clínicos, sem verificação subsequente através de modalidades imagiológicas complementares às radiografias, comprometem a capacidade de generalização da grande maioria dos modelos desenvolvidos.^{93,94} Adicionalmente, destaca-se ainda a potencial ausência de representatividade populacional nos dados de treino que poderá acarretar um enviesamento de desigualdades étnicas, socioeconómicas ou de género que comprometa utilização do modelo em populações diferentes da estudada.¹²⁴ Não obstante, apesar dos avanços recentes, as limitações técnicas e algorítmicas persistem. Mesmo em algoritmos atualmente disponíveis comercialmente, persistem dificuldades na deteção de padrões de fratura incomuns, subtis ou áreas menos exploradas, como na faixa pediátrica. Assim, o grau de incerteza inerente às predições pode induzir clínicos em erro, comprometendo a segurança do doente.¹¹⁷

Ainda nas limitações algorítmicas, muitas vezes os modelos funcionam como “caixas-negras” que fornecem classificações binárias sem demonstração do raciocínio subjacente à decisão, o que dificulta a interpretação e fiabilidade. Ferramentas como *Grad-CAM*, que visam destacar as regiões imagiológicas cruciais para a decisão do modelo, procuram mitigar esta problemática, bem como modelos localizados da região fraturada.¹²⁵ A integração clínica carece ainda muitas vezes de compatibilidade técnica com sistemas preexistentes, como o PACS, bem como a formação contínua dos profissionais de saúde na interpretação dos resultados e na compreensão do seu funcionamento, que representam condições *sine qua non* para uma adoção bem-sucedida.⁹⁶

Numa visão de aplicabilidade para com o doente, pesa a confidencialidade e a proteção de dados que emergem como imperativos éticos e legais. A sensibilidade intrínseca da informação médica adquire novos moldes quando o seu tratamento transcende a esfera exclusivamente humana, devendo por isso haver um esforço incessante para a anonimização rigorosa dos

dados. Despoleta ainda questões legais complexas e de responsabilização jurídica. A atribuição de responsabilidade em casos de erro clínico torna-se um desafio dado a sinergia entre múltiplos intervenientes, que vão desde o programador ao profissional de saúde utilizador, carecendo por isso de uma regulamentação neste sentido.¹²⁶

Com o intuito de responder aos desafios inerentes à aplicação da IA na medicina, a União Europeia tem vindo a estabelecer um enquadramento regulatório que visa garantir que há um desenvolvimento responsável e eticamente orientado destes sistemas. A recém-aprovada Lei Europeia da IA introduz requisitos rigorosos para sistemas classificados como de alto risco, como os utilizados em contextos clínicos, impondo mecanismos de mitigação de riscos, transparência, e fiabilidade dos dados, bem como supervisão humana. Paralelamente, a nova Diretiva Europeia de Responsabilidade pelo Produto reforça os mecanismos de responsabilização por danos decorrentes de falhas nestas tecnologias.¹²⁶

Reconhecendo a importância da necessidade de acesso a dados de qualidade para o treino de modelos acuráveis, a Comissão Europeia prevê, para 2025, a implementação do Espaço Europeu de Dados de Saúde. Este permitirá o acesso transfronteiriço a dados eletrónicos de saúde, garantindo a investigação e inovação tecnológicas neste meio, e, simultaneamente salvaguardando os direitos individuais e a proteção de dados sensíveis. Complementarmente, a iniciativa AICare@EU, coordenada pela Direção-Geral da Saúde e Segurança Alimentar, visa catalisar a adoção segura da IA na prática clínica, promovendo estudos, parcerias internacionais e mecanismos de financiamento orientados à sua integração eficiente nos sistemas de saúde. Este sistema regulatório ambiciona posicionar a Europa na vanguarda da inovação médica, assegurando simultaneamente padrões elevados de ética, equidade e segurança clínica.¹²⁶

No futuro, e tendo em consideração que, mesmo no panorama atual, existe margem para melhoria, é de notar que não foram identificados estudos que explorem a previsão do alinhamento pós-fratura através de modelos de inteligência artificial, o que representaria um contributo valioso para a prática clínica, nomeadamente na estratificação do tratamento e no prognóstico funcional. A combinação de dados clínicos complementares, como antecedentes médicos, mecanismo de trauma e exame físico, poderá igualmente potenciar a precisão diagnóstica dos modelos. Por conseguinte, a aplicação destas tecnologias em países de baixo rendimento demonstra um potencial promissor, considerando a escassez de recursos humanos especializados e a elevada prevalência de fraturas em contextos com acesso limitado a cuidados imagiológicos avançados.¹²⁷

6. Discussão

A presente revisão aborda o estado de arte da aplicação da Inteligência Artificial na detecção de fraturas ósseas. Afigura-se com especial relevância num contexto de crescente pressão sobre os sistemas de saúde, o que por vezes se associa a um risco não menosprezável de erro diagnóstico, emergindo assim como uma ferramenta de apoio de decisões clínicas e com papel na eficiência no diagnóstico imagiológico. Através de modelos baseados em redes neuronais convolucionais, como *EfficientNet*, *YOLOv5* ou *InceptionV3*, tornou-se possível automatizar processos de detecção e classificação de fraturas com níveis de precisão, até então, limitados à capacidade de cognição humana. Esta evolução tem implicações na prática, tanto em sede de triagem urgente como na gestão especializada de fraturas complexas.^{35,69}

No seio desta revisão, foram alvos de especial enfoque regiões anatómicas do punho e da anca por conjugarem uma elevada incidência de fratura com significativo impacto funcional e económico, e por estarem bem estabelecidas na literatura referente ao uso da IA. A nível do punho, modelos como o *RegNetX006* com imagens pré-processadas mostraram resultados notáveis (acurácia de 99,2%), enquanto abordagens segmentadas com *U-Net* permitiram uma análise mais pormenorizada e explicativa das fraturas em estudo.^{66,69} Na anca, o destaque recai sobre modelos de *ensemble* baseados em *EfficientNet B5/B6/B7* que, mesmo com conjuntos de dados reduzidos, atingiram AUC de 0,998 na detecção de fraturas atípicas incompletas, bem como sobre modelos como o *DenseNet-169* acoplado a *FPN*, que demonstraram elevado desempenho na classificação multiclasse e superioridade face a clínicos em contextos de baixa qualidade imagiológica.^{86,92} Estes dados demonstram que, dependendo da arquitetura e do contexto clínico, os modelos podem ser adaptados para tarefas binárias, segmentação, ou mesmo classificação hierarquizada de fraturas complexas. Corroborou-se a importância de dados de treino corretamente anotados, e o papel preponderante da quantidade e qualidade dos mesmos. Debateu-se ainda a importância do estabelecimento de uma *ground truth* verosímil, subindo-se de mais do que um decisor, bem como de modalidades imagiológicas além das radiológicas.^{93,94}

Em termos comparativos diretos com profissionais de saúde, vários estudos revelaram que a IA supera médicos com menor experiência e, em alguns casos, iguala ou ultrapassa ortopedistas e radiologistas. A título de exemplo, na detecção de fraturas do escafóide, o modelo identificou 90,9% das fraturas ocultas, superando os ortopedistas em estudo.⁴⁰ No caso da anca, modelos como o *PelviXNet* e variantes baseadas em *VGG-16* e *Inception* demonstraram desempenho superior a de médicos do serviço de urgência e internos, e desempenho comparável ao de especialistas em Ortopedia e Radiologia.⁹⁰

Contudo, impera o reconhecimento de limitações inerentes. Muitos estudos apresentam amostras reduzidas e não representativas, além de recorrerem a *ground truths* heterogêneas, frequentemente baseadas em relatórios clínicos e não confirmadas por TC ou RM.⁹⁰ A ausência de validação externa; a dependência de imagens de qualidade elevada; a exclusão de projeções laterais e a utilização de *datasets* selecionadas, com prevalência artificialmente aumentadas, comprometem a translação dos modelos para contextos clínicos reais.⁹⁶ Acresce ainda que, nos modelos comerciais como o BoneView©, embora a performance seja aceitável na maioria das fraturas comuns, a sua eficácia diminui em fraturas pediátricas específicas e em padrões mais subtis, como as fraturas em trampolim ou as Salter-Harris II¹¹⁷ Por fim, a interpretabilidade, responsabilidade diagnóstica, e integração prática em sistemas de saúde continuam a ser barreiras relevantes que exigem desenvolvimento normativo e validação em estudos prospectivos.¹²⁶

7. Conclusão

Em derradeira análise, torna-se evidente que, apesar da heterogeneidade metodológica e da multiplicidade de variáveis analisadas nos estudos supramencionados, está patente o potencial de uma simbiose entre a IA e o ser humano. Esta colaboração, não deve descorar as responsabilidades clínicas intrínsecas ao médico, uma vez que os dados apontam para a IA como uma solução complementar, e nunca como um substituo humano. Promete aliar-se como uma ferramenta que procura dar resposta a desafios como a sobrecarga dos serviços de saúde, os erros diagnósticos e até os custos associados ao processo clínico.

Dada a natureza dinâmica e em constante mutação deste domínio, torna-se essencial manter revisões regulares e atualizadas, de modo a acompanhar os avanços tecnológicos, bem como garantir a relevância clínica dos sistemas de ponta disponíveis.

Para além das questões algorítmicas, técnicas e médico-legais que inevitavelmente se levantam, importa não esquecer que a medicina é feita de pessoas e para pessoas. A integração da IA no fluxo hospitalar possui, simultaneamente, um potencial catalisador na valorização da relação médico-doente, na medida em que permite maior disponibilidade para a interação humana, mas acarreta concomitantemente um risco de distanciamento, caso a sua utilização não seja corretamente contextualizada. A aceitação por parte dos doentes dependerá de múltiplos fatores, entre os quais se destacam a comunicação clara dos benefícios e da segurança e limitações desta ferramenta, bem como a manutenção de um cuidado centrado no doente.

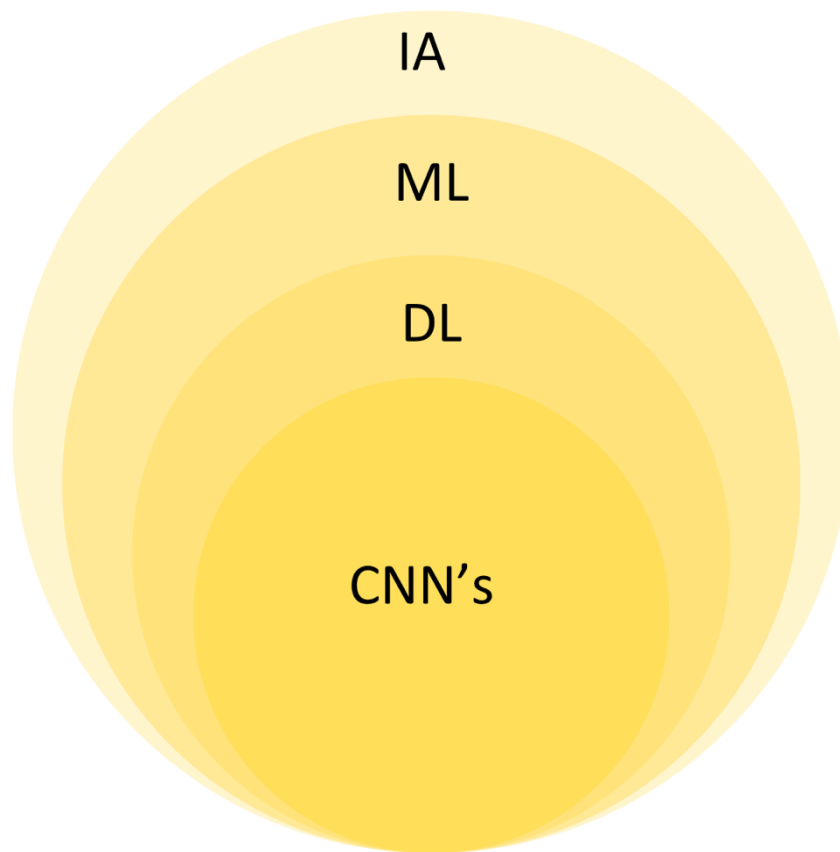


Figura 1- Esquematização dos subtermos de IA

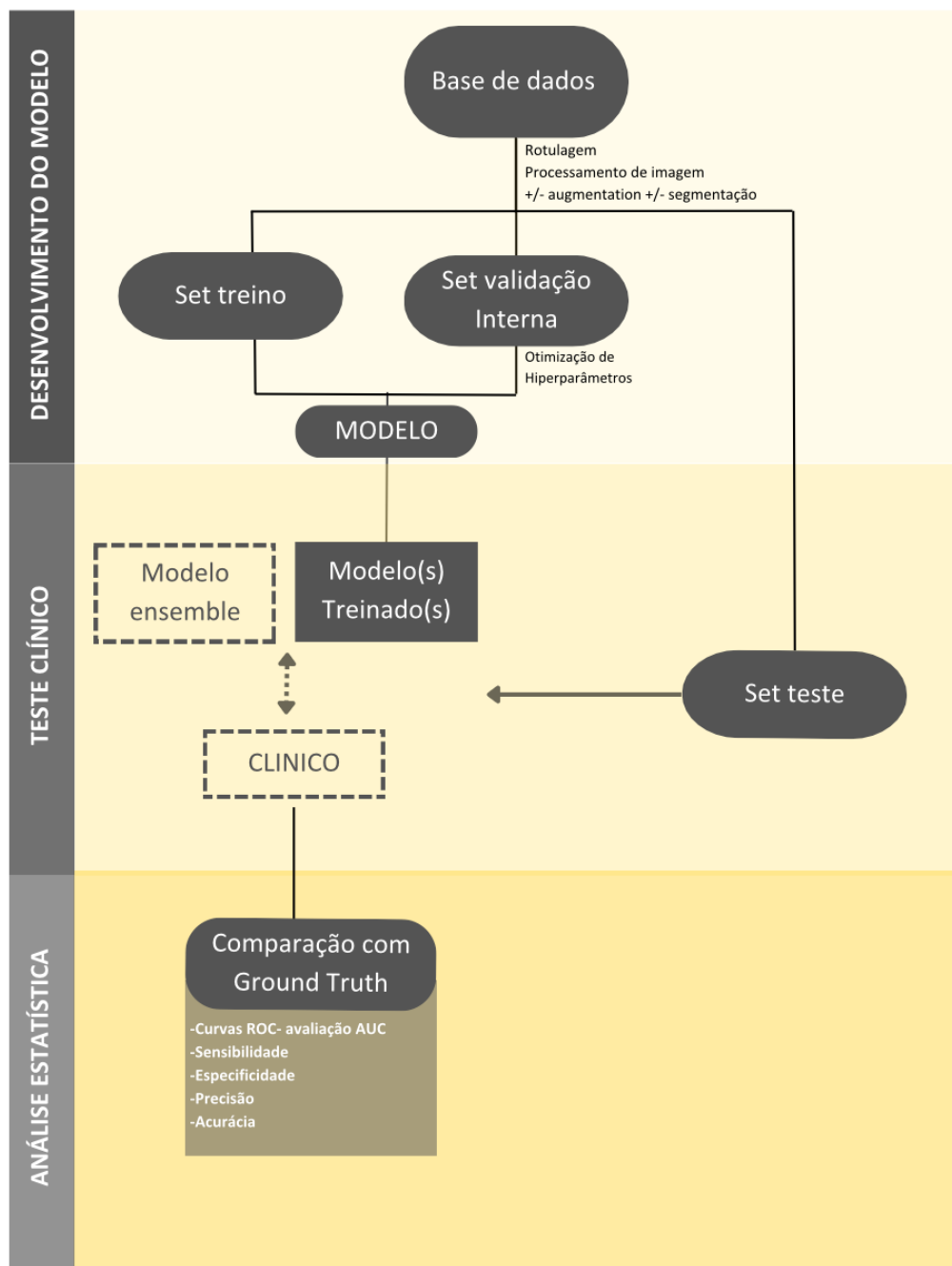
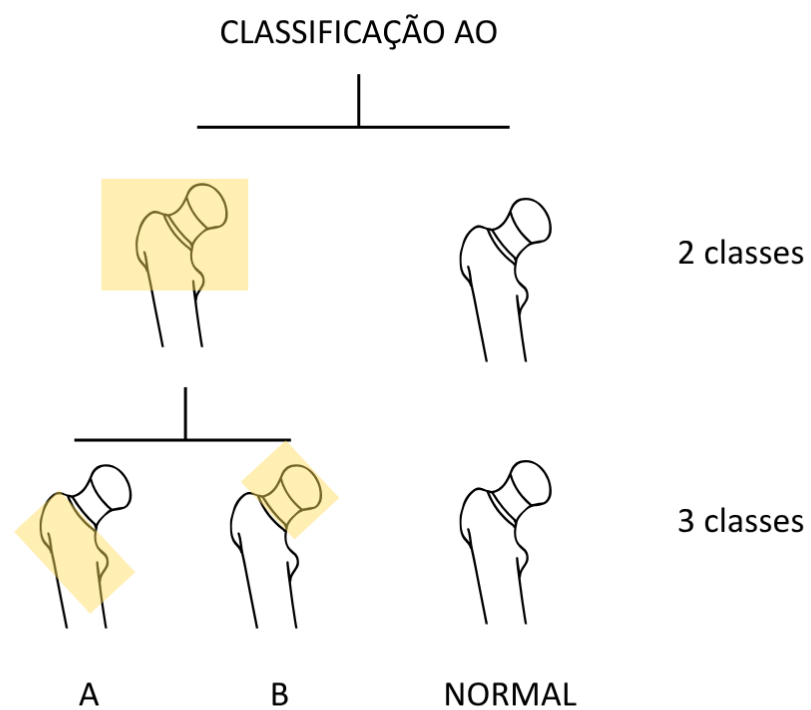


Figura 2- Processo de treino dos modelos de IA aplicados na detecção de fraturas

$$\text{IoU} = \frac{\text{área de sobreposição}}{\text{área de união}}$$

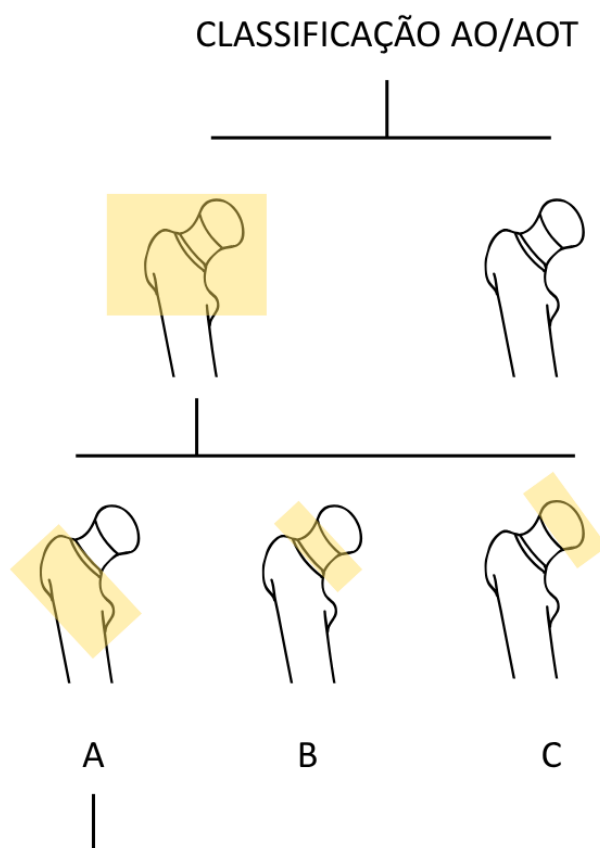

Adaptado de Gan et al.

Figura 3- Cálculo de IoU



Adaptado de Jiménez-Sánchez et al.

Figura 4- Classificação AO adaptada em Jiménez- Sánchez et al.



A1-Fraturas pertrocantéricas simples

A2- Fraturas pertrocantéricas multifragmentares
com parede lateral incompetente

A3-Fraturas intertrocantéricas

Adaptado de Tanzi et al.

Figura 5- Classificação AO/AOT

Tabela I – Estudos de modelos de IA aplicados à detecção de fraturas da região do punho

1º Autor e Ano	Tipo de Imagem Médica e Visão	Metodologia / Ground Truth / Modelo de IA	Tamanho do Treino / Validação / Teste	Output do Modelo	Resultados
Erdaş et al., 2023	Radiografias punho- AP e lateral	Metodologia: DL na detecção de fraturas do cúbito e rádio com diferentes níveis de processamento de imagem ——— <i>Ground Truth</i> : Avaliação por especialistas ——— Modelo IA: <i>RegNetX006</i> , <i>EfficientNet B0</i> , <i>InceptionResNetV2</i>	Treino: 8863 (4480 imagens com fratura, 4383 sem fratura) ——— Validação: Não especificado ——— Teste: Não especificado	Classificação binária (fratura/não fratura)	Imagens brutas (acurácia; sensibilidade; especificidade) <i>RegNetX006</i> – 97,8% 98,7% 95,4% <i>EfficientNet B0</i> – 94,0% 97,2% 86,0% <i>InceptionResNetV2</i> – 91,6% 89,5% 96,7% Imagens com remoção apenas do fundo (acurácia; sensibilidade; especificidade) <i>RegNetX006</i> – 98,2% 98,8% 97,7% <i>EfficientNet B0</i> – 97,6% 98,9% 96,4% <i>InceptionResNetV2</i> – 93,9% 94,1% 93,7% Imagens com remoção de tecidos não ósseos (acurácia; sensibilidade; especificidade) <i>RegNetX006</i> – 99,2% 99,7% 98,6% <i>EfficientNet B0</i> – 97,3% 96,2% 98,4% <i>InceptionResNetV2</i> – 94,7% 94,9% 94,5%
Anttila et al., 2023	Radiografias punho- AP e lateral	Metodologia: Segmentação de FRD baseada em U-Net ——— <i>Ground Truth</i> : Avaliação por ortopedistas ——— Modelo IA: U-Net (25 camadas, segmentação)	Treino: 3399 imagens ——— Validação: Não especificado ——— Teste: 386 imagens	Segmentação com mapa de confiança por pixel para presença de fratura	Todas as radiografias (acurácia sensibilidade especificidade) 91,0% 92,0% 88,0% Projeção lateral apenas (acurácia sensibilidade especificidade) 89,0% 90,0% 87,0% Projeção AP apenas (acurácia sensibilidade especificidade) 93,0% 94,0% 90,0% Média da confiança lateral + PA (acurácia sensibilidade especificidade) 94,0% 95,0% 91,0%
Min et al., 2023	Radiografia do punho AP	Metodologia: DL em duas etapas para classificação em FRD em intra e extra-articulares ——— <i>Ground Truth</i> : Avaliação por ortopedistas ——— Modelo IA: <i>YOLOv5</i> e <i>EfficientNetB3</i> (<i>ensemble</i>)	Treino: 334 imagens com fraturas (218 intra-articulares e 74 extra-articulares) ——— Validação: Não especificado ——— Teste: 57 imagens com fratura (42 intra-articulares e 15 extra-articulares)	Classificação binária (fratura intra-articular vs. extra-articular)	AUC: 0.82 Acurácia: 81,0% Sensibilidade: 83,0% Especificidade: 73,0%

Yoon et al., 2021	Radiografia do escafoide (PA e <i>scaphoid view</i>)	Metodologia: <i>EfficientNetB3</i> para classificar fraturas do escafoide em radiografias, utilizando um modelo de dois estadios para fraturas aparentes e ocultas ——— Ground Truth: Fraturas confirmadas por radiologistas através de TC ou RM ——— Modelo IA: Cascade R-CNN e 2 modelos de <i>EfficientNetB3</i>	Treino: 8356 imagens ——— Validação: 1177 imagens ——— Teste: 2305 imagens	Classificação binária (fratura/não fratura)	AUC: 0,96 (fraturas aparentes) e 0,81(fraturas ocultas) Sensibilidade: 87,1% (aparentes), 79,0% (ocultas) Especificidade: 92,1% (aparentes), 71,6% (ocultas)
Lee et al., 2023	Radiografia do punho (AP, lateral e oblíqua)	Metodologia: DL para a detecção FRD, apófise estilóide cubital e escafoide, de forma independente e posteriormente combinado com clinico que analisou as imagens sem IA inicialmente, e após um período de <i>washout</i> de 3 semanas voltou a analisar com auxilio de IA ——— <i>Ground Truth</i> : Avaliação por ortopedista e radiologista ——— Modelo IA: <i>RetinaNet</i> , <i>DeepLab v3</i> e <i>NasNet</i>	Treino: 4432 imagens ——— Validação: 20% dos dados de treino ——— Teste: 1186 imagens	Classificação binária (fratura/não fratura)	Desempenhos (Sensibilidade Especificidade Acurácia AUC) IA Rádio Distal: 97,2% 83,2% 87,2% 0,90 IA Apófise estilóide cubital: 97,7% 87,3% 89,6% 0,93 IA Escafoide: 87,0% 74,0% 74,0% 0,81 Radiologistas (sem IA → com IA): Rádio Distal: 90,0–94,0% → 95,0–97,0% 95,0–97,0% → 94,0–97,0% 94,0–96,0% → 94,0–97,0% 0,93–0,96 → 0,94–0,97 Apófise estilóide cubital: 71,0–97,0% → 86,0–98,0% 97,0–99,0% → 95,0–97,0% 93,0–97,0% → 94,0–97,0% 0,85–0,97 → 0,91–0,97 Escafoide: 47,0–50,0% → 66,0–72,0% 96,0–99,0% → 94,0–98,0% 88,0–98,0% → 93,0–98,0% 0,71–0,75 → 0,80–0,85

Gan et al., 2019	Radiografia AP do punho	Metodologia: DL na detecção de ROI e classificação de FRD em comparação com clínicos ——— Ground Truth: Avaliação manual por ortopedistas e verificação com TC em casos controversos ——— Modelo IA: <i>Faster R-CNN</i>, <i>Inception-v4</i>	Treino: 2.040 imagens (1.341 com fratura, 699 sem) ——— Validação: 15% dos dados de treino usados para validação ——— Teste: 300 imagens (150 com fratura, 150 sem)	Probabilidade de fratura (0 a 1) com <i>Bounding boxes</i>	Desempenho <i>Faster R-CNN</i>: Média do IOU: 0,87 Comparação de desempenho – IA ensemble vs. médicos: Acurácia: IA 93,0% Ortopedistas 94,0% Radiologistas 84,0% Sensibilidade: IA 90,0% Ortopedistas 93,0% Radiologistas 81,0% Especificidade: IA 96,0% Ortopedistas 95,0% Radiologistas 87,0%
Suzuki et al., 2021	Radiografia (AP e lateral do punho)	Metodologia: DL na classificação de FRD ——— Ground Truth: Diagnóstico confirmado por dois ortopedistas ——— Modelo IA: <i>EfficientNet (B2, B3, B4 e B5)</i>	Treino: 755 imagens (AP), 578 imagens (lateral) ——— Validação: Não especificado ——— Teste: 150 imagens	Classificação binária (fratura/não fratura)	Desempenho do modelo Ensemble (<i>EfficientNet-B2 e B4</i>) vs Ortopedistas Acurácia: IA 99,3% Ortopedistas: 94,7–97,3% Sensibilidade: IA 98,7% Ortopedistas: 96,0% Especificidade: IA 100% Ortopedistas: 93,3–98,7%
Zhang et al., 2023	Radiografia de punho (AP e lateral)	Metodologia: DL na classificação de FRD ——— Ground Truth: Diagnóstico confirmado por três radiologistas ——— Modelo IA: <i>Faster R-CNN</i>, <i>RetinaNet</i>, <i>Cascade R-CNN</i>	Treino: 2.296 imagens AP e 2.283 laterais ——— Validação: 491 AP e 488 laterais ——— Teste: 489 AP e 489 laterais	Bounding boxes para fraturas	Desempenho do modelo vs clínicos utilizando imagens AP e Laterais combinadas Acurácia: IA: 97,6% Ortopedistas: 93,9% Radiologistas: 92,9% Sensibilidade: IA: 98,4% Ortopedistas: 95,9% Radiologistas: 95,4% Especificidade: IA: 96,7% Ortopedistas: 91,8% Radiologistas: 90,5%

Langerhuizen et al., 2020	Radiografia do escafoide (PA, lateral, oblíqua, ulnar deviation)	Metodologia: CNN pré-treinada (<i>VGG16</i>) para detecção de fraturas do escafoide em radiografias, utilizando <i>data augmentation</i> e validação cruzada — — — Ground Truth: Fraturas confirmadas por radiologistas com recurso a TC ou RM — — — Modelo IA: <i>VGG16</i>	Treino: 180 imagens — — — Validação: 20 imagens — — — Teste: 100 imagens	Classificação binária (fratura/não fratura)	Desempenho de modelo vs Ortopedistas Especificidade: Ortopedistas: 93,0% IA: 60,0% Acurácia: Ortopedistas: 84,0% IA: 72,0% Sensibilidade: Ortopedistas: 76,0% IA: 84,0%
----------------------------------	--	--	--	---	---

Tabela II – Estudos de modelos de IA aplicados à detecção de fraturas da região da anca

1º Autor e Ano	Tipo de Imagem Médica e Visão	Metodologia / Ground Truth / Modelo de IA	Tamanho do Treino / Validação / Teste	Output do Modelo	Resultados
Kim et al., 2023	Radiografia AP do fêmur	Metodologia: <i>Transfer learning</i> com <i>ensemble</i> de CNNs, aplicação de filtro Sobel, data augmentation ---- Ground Truth: Avaliação por radiologista musculoesquelético e cirurgião ortopédico com 15 anos de experiência ---- Modelo IA: <i>EfficientNet B5/B6/B7, DenseNet121, MobileNet V1/V2, ensemble</i> dos 3 e dos 5 melhores modelos	Treino: 630 imagens (570 radiografias normais e 60 fraturas) ---- Validação: 380 imagens (380 normais e 40 fraturas) ---- Teste: Teste externo com 40 imagens (20 normais + 20 fraturas)	Classificação binária (fratura/não fratura)	Ensemble Top 3: Acurácia: 99,3% Sensibilidade 95,0% AUC 0,998 Ensemble Top 5: Acurácia: 99,1% Sensibilidade 92,5% AUC 0,997
Alzaid et al., 2022	Radiografia do fêmur (AP e lateral)	Metodologia: Classificação e detecção de fraturas periprotéticas do fêmur usando CNNs e modelos de detecção de objetos ---- Ground Truth: Anotação por dois especialistas clínicos segundo o Vancouver Classification System (VCS) ---- Modelo IA: <i>Densenet161, Resnet50, Inception, VGG, Faster R-CNN, RetinaNet</i>	Treino: 1908 imagens no modelo binário, 912 imagens no modelo multiclasse ---- Validação: 636 imagens no modelo binário, 304 imagens no modelo multiclasse ---- Teste: Não Especificado	Classificação binária (fratura/normal) e multi-classe (Tipo A, B, C e normal), localização da fratura	Classificação binária e Localização de FPFs -Faster R-CNN: Precisão: 80,0%; Sensibilidade: 98,0%; Acurácia: 78,0% -RetinaNet: Precisão: 31,0%; Sensibilidade: 97,0%; Acurácia: 31,0% Classificação binária -Melhor acurácia: <i>DenseNet</i> com 95,0% -Melhor sensibilidade: <i>Inception e ResNet50</i> com 96,0% -Melhor especificidade: <i>DenseNet</i> com 97,0% Classificação multiclasse de FPFs -Melhor acurácia média para os 4 tipos: <i>ResNet50</i> com 90,0% -Melhor sensibilidade média para os 4 tipos : <i>Inception e DenseNet</i> com 75,0% -Melhor especificidade média para os 4 tipos: <i>Inception e ResNet50</i> com 93,0%
Ouyang et al., 2023	Radiografia pélvica, visão anteroposterior	Metodologia: Aplicação de <i>Design Thinking</i> no desenvolvimento de um modelo de DL para detecção de fraturas do fêmur, utilizando entrevistas clínicas ---- Ground Truth: Relatórios que acompanharam as radiografias no sistema PACs---- Modelo IA: <i>Xception CNN</i>	Treino: 80% do conjunto de dados ---- Validação: Não Especificado ---- Teste: 100 imagens (25 fraturas do colo femoral, 25 intertrocantericas, 50 normais)	Classificação binária (fratura/normal)	Desempenho do modelo antes e após Design Thinking -Sensibilidade: 97,0% → 97,0% -Especificidade: 84,0% → 93,0% -Acurácia: 91,0% → 95,0%

Urakawa et al., 2018	Radiografia pélvica, visão AP	Metodologia: Estudo retrospectivo comparativo entre IA e ortopedistas para detecção de fraturas intertrocantericas em Radiografia pélvica, que foram recortadas de modo a exibir apenas os fêmures proximais Ground Truth: Diagnóstico confirmado pelo registo do tratamento cirúrgico e revisão de imagens por ortopedistas Modelo IA: <i>VGG_16</i>	Treino: 2678 imagens (1408 fraturadas, 1270 não-fraturadas) Validação: 334 imagens (185 fraturadas, 149 não-fraturadas) Teste: 334 imagens (180 fraturadas, 154 não-fraturadas)	Classificação binária (fraturado/não-fraturado)	Desempenho da IA Acurácia: 95,5% Sensibilidade: 93,9% Especificidade: 97,4% Desempenho de Ortopedistas Acurácia: 92,2% Sensibilidade: 88,3% Especificidade: 96,8%
Cheng C.T et al., 2021	Radiografia pélvica, visão AP	Metodologia: DL treinada com radiografias com anotação baseada em ponto-a-ponto por especialistas e comparação com 22 clinicos de diversas especialidades Ground Truth: Anotação por especialistas (radiologistas, ortopedistas, cirurgiões de trauma) Modelo IA: <i>PelviXNet (DenseNet-169 e FPN)</i>	Treino: 5204 imagens Validação: Não Especificado Teste: 1888 imagens	Classificação binária (fraturado/não-fraturado)	Fraturas da anca (acurácia, sensibilidade, especificidade) -<i>PelviXNet</i>: 99,5%, 100%, 99% -Médicos de Urgência: 95%, 97,5%, 92,5% -Média (Radiologistas + ortopedistas): 97,4%, 99,5%, 95,3% -Internos: 94,9%, 95%, 94,8% Fraturas pélvicas (acurácia, sensibilidade, especificidade) -<i>PelviXNet</i>: 94,5% 92,0% 97,0% -Médicos do Serviço de Urgência: 90,3% 85,0% 95,5% -Média (Radiologistas + ortopedistas): 94,0% 89,5% 99,5% -Internos: 89,3% 84,4% 94,1%
Kroguet et al., 2020	Radiografia pélvica, visão AP e lateral	Metodologia: Estudo retrospectivo sobre detecção e subclassificação automática de fraturas da anca DL Ground Truth: Revisão por ortopedistas com confirmação por TC, RM e imagens pós-operatórias Modelo IA: <i>ResNet-50-FPN</i> para detecção automática de <i>bounding boxes</i> <i>DenseNet-169</i> para classificação	Treino: 1849 imagens Validação: 739 imagens Teste: 438 imagens	Classificação binária (fraturado/não fraturado) e classificação multiclasse (Colo do fémur deslocada/não deslocada, intertrocanterica, fixação interna prévia, artroplastia prévia, normal)	Comparação IA e Médico num sub-teste de 100 imagens: Médicos especialistas (imagens baixa resolução sem IA → com IA imagens alta resolução, sem IA) -Acurácia binária: 88,6% → 95,5% 93,5% -Acurácia multiclasse: 83,4% → 92,7% 89,9% Residentes (Baixa resolução → com IA Alta resolução, sem IA): -Acurácia binária: 83,9% → 91,0% 85,6% -Acurácia multiclasse: 76,1% → 88,3% 78,5% Modelo de IA -Acurácia binária: 95,8% -Acurácia multiclasse: 92,8%

Tsubasa Mawatari et al., 2020	Radiografia Pélvica, visão AP	Metodologia: Detecção de fraturas do fémur proximal utilizando CNNs e avaliação do impacto no desempenho dos radiologistas ——— Ground Truth: Anotação manual por radiologistas com referência a TC e RM. Classificadas em: 1.Fratura evidente na radiografia. 2.Fratura possível na radiografia. 3.Fratura obscura, mas detetável na radiografia 4.Fratura diagnosticada apenas por TC 5.Fratura diagnosticada apenas por RM ——— Modelo IA: <i>GoogLeNet (Inception V1)</i>	Treino: 302 imagens ——— Validação: Não especificado ——— Teste: 50 imagens	Classificação binária (fraturado/não-fraturado)	Desempenho IA Sensibilidade: 88,0% Especificidade 72,0% AUC : 0,905 Desempenho clínicos Sensibilidade média:78,3% Especificidade média: 72,6% AUC média de todos os leitores (sem IA→ com IA): 0,832 → 0,876
Jiménez-Sánchez et al., 2020	Radiografia Pélvica, visão AP	Metodologia: DL para deteção e classificação automática de fraturas do fémur proximal segundo a classificação AO, em comparação com 3 clínicos. ——— Ground Truth: <i>Bounding boxes</i> e classificação anotadas por radiologistas e ortopedistas ——— Modelo IA: <i>AlexNet</i> (localização) <i>ResNet-50</i> (classificação)	Treino: 943 imagens ——— Validação: 135 imagens ——— Teste: 269 imagens	Classificação de fratura (tipo A, B ou sem fratura)	Desempenho classificação binária -IA(radiografia completa, ROI manual, localização automática): Acurácia: 83,0%; 93,0%; 93,0% Sensibilidade: 83% , 94% , 93% -Média dos 3 clínicos: Acurácia: 92,0% Sensibilidade: 92,0% Desempenho classificação em 3 classes Fratura Tipo A: -IA(radiografia completa, ROI manual, localização automática): Acurácia: 86,0%; 91,0%; 89,0% Sensibilidade: 67,0%; 75,0%; 73,0 -Média dos 3 clínicos: Acurácia: 92,0% Sensibilidade: 83,0% Fratura Tipo B: -IA(radiografia completa, ROI manual, localização automática): Acurácia: 87%, 91%, 87% Sensibilidade: 83%, 88%, 90% -Média dos 3 clínicos: Acurácia: 89,0% Sensibilidade: 92,0% Normal: -IA(radiografia completa, ROI manual, localização automática): Acurácia: 94,0%; 91,0%; 94,0% Sensibilidade: 95,0%, 97,0%, 92,0% -Média dos 3 clínicos: Acurácia: 93,0% Sensibilidade: 86,0%

Leonardo Tanzi et al., 2020	Radiografia Pélvica, visão AP	Metodologia: Classificação hierárquica de fraturas do fémur proximal com CNNs em cascata e comparação com desempenho clínico ——— Ground Truth: Anotações por ortopedistas e radiologistas ——— Modelo IA: <i>InceptionV3</i> com abordagem <i>multistage</i>	Treino: 1962 imagens ——— Validação: 491 imagens ——— Teste: 491 imagens	Classificação de fraturas do fémur proximal segundo AO/OTA (tipo A1,A2,A3,B e Normal)	Desempenho IA – Arquitetura <i>Multistage</i>: Acurácia para 3 classes (A, B, normal): 87,0% Acurácia para 5 classes (A1, A2, A3, B, normal): 80%→ Após adição de rede binária para distinguir entre A1 e A2: 81% Desempenho Médico(Acurácia sem IA → com IA) Binária: -Radiologista: 99,0% -Estudante: 95,0% Fratura tipo A vs. B: -Radiologista: 95,0% -Estudante: 89,0% Fraturas A1 vs. A2 vs. A3: -Radiologista: 73,0% → 86,0% -Estudante: 68,0% → 83,0%
------------------------------------	-------------------------------	--	---	---	--

Tabela III– Estudos de modelos de IA aplicados à detecção de fraturas pediátricas

1ºAutor e Ano	Tipo de Imagem Médica e Visão	Metodologia / Ground Truth / Modelo de IA	Tamanho do Treino / Validação / Teste	Output do Modelo	Resultados
Zech et al., 2023	Radiografia de punho (projeção AP)	Metodologia: Estudo retrospectivo de performance com leitura por 4 internos (2 de pediatria e 2 de radiologia) com e sem IA (intervalo de 3–12 semanas) ---- Ground Truth: Relatório original feito por radiologista e segmentação por interno ---- Modelo IA: <i>Faster R-CNN</i>	Treino: 229 imagens ---- Validação: 41 imagens ---- Teste: 125 imagens	Caixa delimitadora de fratura + classificação binária (com fratura / sem fratura)	Desempenho da IA: AUC: 0,92 Acurácia global: 88,0% Sensibilidade: 88,0% Especificidade: 89,0% Fraturas em buckle: 90,0% detectadas corretamente Desempenho dos Internos (Sem → Com IA): Acurácia global: 80,0% → 93,0% Acurácia em fraturas de buckle: 69,0% → 92,0% Sensibilidade: 91,0% → 96,0%
Zech et al., 2024	Radiografias do membro superior (cotovelo, mão/dedo, úmero/ombro/clavícula, punho/antebraço)	Metodologia: Estudo de várias regiões do membro superior, com leitura por 2 radiologistas especialistas, 3 internos de radiologia e 2 de pediatria com e sem IA (3–4 semanas de intervalo) ---- Ground Truth: Consenso entre dois radiologistas musculoesqueléticos ---- Modelo IA: <i>Faster R-CNN X101-FPN</i>	Sem treino	Caixa delimitadora de fratura + classificação binária (com fratura / sem fratura)	Desempenho IA: AUC: 0,89 Sensibilidade: 90,0% Especificidade: 88,3% Desempenho Médicos (sem IA→com IA) Especialistas: AUC: 0,88 → 0,89 Sensibilidade: 84,2% → 85,8% Especificidade: 89,2% → 92,1% Internos de Radiologia: AUC: 0,77 → 0,88 Sensibilidade: 78,1% → 88,3% Especificidade: 75,6% → 86,9% Internos de Pediatria: AUC: 0,71 → 0,84 Sensibilidade: 65,8% → 83,3% Especificidade: 75,4% → 85,4%
Ghosh et al., 2023	Radiografia torácica frontal	Metodologia: Estudo retrospectivo com treino e teste de CNNs em radiografias pediátricas para detecção de fraturas costais ---- Ground Truth: Relatórios radiológicos revistos por radiologista ---- Modelo IA: <i>ResNet-18, ResNet-50</i> (pré-treinados no ImageNet)	Treino: 853 imagens ---- Validação: 197 imagens ---- Teste: 261 imagens	Classificação binária: fratura / sem fratura	Sensibilidade Radiologista 1: 88,0% Radiologista 2: 72,0% <i>ResNet-50</i> : 82,0% <i>DenseNet-121</i> : 76,0% Especificidade Radiologista 1: 75,0% Radiologista 2: 100% <i>ResNet-50</i> : 75,0% <i>DenseNet-121</i> : 79,0%

Binh et al., 2024	Radiografia do punho, visão AP e lateral	Metodologia: Estudo de modelo de detecção multiclases com base na classificação AO/OTA, utilizando CNN <i>YOLOv4</i> e comparação com ortopedista e radiologista ———	Treino: 1447 imagens ——— Validação: 362 imagens ———	Detecção e classificação de fraturas (FRM, FUM, FRE, FUE)	Ulnar Metaphysis Fracture (FUM) Modelo: 71,4% sensibilidade 94,5% especificidade AUC 0,83 Ortopedista: 57,1% sensibilidade 91,7% especificidade AUC 0,74 Radiologista: 92,8% sensibilidade 97,2% especificidade AUC 0,95 Ulnar Epiphysis Fracture (FUE) Modelo: 89,4% sensibilidade 98,5% especificidade AUC 0,94 Ortopedista: 84,2% sensibilidade 98,5% especificidade AUC 0,91 Radiologista: 89,4% sensibilidade 98,5% especificidade AUC 0,94 Radial Epiphysis Fracture (FRE) Modelo: 88,2% sensibilidade 97,1% especificidade AUC 0,93 Ortopedista: 76,4% sensibilidade 95,7% especificidade AUC 0,86 Radiologista: 76,4% sensibilidade 98,5% especificidade AUC 0,88 Radial Metaphysis Fracture (FRM) Modelo: 86,4% sensibilidade 88,0% especificidade AUC 0,87 Ortopedista: 81,1% sensibilidade 80,0% especificidade AUC 0,81 Radiologista: 94,5% sensibilidade 90,0% especificidade AUC 0,92
		Ground Truth: Anotações manuais com bounding boxes por ortopedistas e radiologista, usando a classificação AO/OTA como referência ——— Modelo IA: <i>YOLOv4</i>	Teste: 300 imagens		
Knight et al., 2023	Ecografia 2D e 3D do punho	Metodologia: Estudo diagnóstico prospectivo com recurso a imagens ecográficas. Leitura e avaliação cruzada entre médicos e IA ——— Ground Truth: Radiografia reavaliada por radiologista musculoesquelético pediátrico para a decisão do diagnóstico final. Utilizador de ecógrafo experiente para a identificação de fraturas na ecografia ——— Modelo IA: <i>ResNet34</i> e <i>DenseNet121</i>	Treino: 16 865 imagens de varredura (2D), 15 882 (3D) ——— Validação: 4 215 imagens de varredura (2D), 3 787 (3D) ——— Teste: 3 822 imagens de varredura(2D), 4 034 (3D)	Classificação binária (fratura vs. normal) com segmentação	Desempenho Modelo IA ResNet34 (Ecografia 2D Ecografia 3D) -Acurácia: 0,89 1,00 -Sensibilidade (SN): 0,82 1,00 -Especificidade (SP): 1,00 1,00 -AUC(2DUS): 0,93 DenseNet121 (Ecografia 2D Ecografia 3D) -Acurácia: 0,94 0,94 -Sensibilidade (SN): 0,91 1,00 -Especificidade (SP): 1,00 0,86 - AUC(3D): 0,91 Desempenho Clínico – Intervalos Sensibilidade -Ecografia 2D: 0,62 – 0,97 -Ecografia 3D: 0,63 – 0,98 Especificidade -Ecografia 2D: 0,82 – 0,94 -Ecografia 3D: 0,91 – 0,95

Referências Bibliográficas

1. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Family Med Prim Care* 2019;8(7):2328-2331. (In eng). DOI: 10.4103/jfmpc.jfmpc_440_19.
2. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism - Clinical and Experimental* 2017;69:S36-S40. DOI: 10.1016/j.metabol.2017.01.011.
3. Cornet G. Chapter 4. Robot companions and ethics: A pragmatic approach of ethical design. *Journal International de Bioéthique* 2013;Vol. 24(4):49-58. (In en). DOI: 10.3917/jib.243.0049.
4. Cabitza F, Locoro A, Banfi G. Machine Learning in Orthopedics: A Literature Review. *Front Bioeng Biotechnol* 2018;6:75. (In eng). DOI: 10.3389/fbioe.2018.00075.
5. Fayed AM, Mansur NSB, de Carvalho KA, Behrens A, D'Hooghe P, de Cesar Netto C. Artificial intelligence and ChatGPT in Orthopaedics and sports medicine. *J Exp Orthop* 2023;10(1):74. (In eng). DOI: 10.1186/s40634-023-00642-8.
6. Landin LA. Fracture patterns in children. Analysis of 8,682 fractures with special reference to incidence, etiology and secular changes in a Swedish urban population 1950-1979. *Acta Orthop Scand Suppl* 1983;202:1-109. (In eng).
7. Rupp M, Walter N, Pfeifer C, et al. The Incidence of Fractures Among the Adult Population of Germany—an Analysis From 2009 through 2019. *Dtsch Arztebl Int* 2021;118(40):665-669. (In eng). DOI: 10.3238/arztebl.m2021.0238.
8. Portugal P. Censos 2021 - Resultados Definitivos. Fundação Francisco Manuel dos Santos. [Acedido em 2024 set 15]. Disponível em: <https://encurtador.com.br/X92Da>. [Acedido em 2024 set 15]. Disponível em: <https://encurtador.com.br/X92Da>.
9. Brazell CJ, Carry PM, Holmes KS, Salton RL, Hadley-Miller N, Georgopoulos G. Pediatric and Adult Fracture Incidence: A Decreasing Trend With Increasing Hospital Admissions. *Orthopedics* 2023;46(6):e369-e375. (In eng). DOI: 10.3928/01477447-20230329-01.
10. Hallas P, Ellingsen T. Errors in fracture diagnoses in the emergency department--characteristics of patients and diurnal variation. *BMC Emerg Med* 2006;6:4. (In eng). DOI: 10.1186/1471-227x-6-4.
11. Wei CJ, Tsai WC, Tiu CM, Wu HT, Chiou HJ, Chang CY. Systematic analysis of missed extremity fractures in emergency radiology. *Acta Radiol* 2006;47(7):710-7. (In eng). DOI: 10.1080/02841850600806340.
12. Whang JS, Baker SR, Patel R, Luk L, Castro A, 3rd. The causes of medical malpractice suits against radiologists in the United States. *Radiology* 2013;266(2):548-54. (In eng). DOI: 10.1148/radiol.12111119.
13. Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci U S A* 2018;115(45):11591-11596. (In eng). DOI: 10.1073/pnas.1806905115.
14. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet Healthy Longev* 2021;2(9):e580-e592. (In eng). DOI: 10.1016/s2666-7568(21)00172-0.
15. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointestinal Endoscopy* 2020;92(4):807-812. DOI: 10.1016/j.gie.2020.06.040.
16. Weiss S, Kulikowski CA, Safir A. Glaucoma consultation by computer. *Computers in Biology and Medicine* 1978;8(1):25-40. DOI: [https://doi.org/10.1016/0010-4825\(78\)90011-2](https://doi.org/10.1016/0010-4825(78)90011-2).
17. Kulikowski CA. Beginnings of Artificial Intelligence in Medicine (AIM): Computational Artifice Assisting Scientific Inquiry and Clinical Art - with Reflections on Present AIM Challenges. *Yearb Med Inform* 2019;28(1):249-256. (In eng). DOI: 10.1055/s-0039-1677895.

18. Bohr A, Memarzadeh K. Chapter 1 - Current healthcare, big data, and machine learning. In: Bohr A, Memarzadeh K, eds. *Artificial Intelligence in Healthcare*: Academic Press; 2020:1-24.
19. An Q, Rahman S, Zhou J, Kang JJ. A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges. *Sensors (Basel)* 2023;23(9) (In eng). DOI: 10.3390/s23094178.
20. Turner OC, Aeffner F, Bangari DS, et al. Society of Toxicologic Pathology Digital Pathology and Image Analysis Special Interest Group Article*: Opinion on the Application of Artificial Intelligence and Machine Learning to Digital Toxicologic Pathology. *Toxicol Pathol* 2020;48(2):277-294. (In eng). DOI: 10.1177/0192623319881401.
21. Beam AL, Kohane IS. Translating Artificial Intelligence Into Clinical Care. *Jama* 2016;316(22):2368-2369. (In eng). DOI: 10.1001/jama.2016.17217.
22. Tajmir SH, Alkasab TK. Toward Augmented Radiologists: Changes in Radiology Education in the Era of Machine Learning and Artificial Intelligence. *Acad Radiol* 2018;25(6):747-750. (In eng). DOI: 10.1016/j.acra.2018.03.007.
23. Sohn K, Berthelot D, Carlini N, et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems* 2020;33:596-608.
24. Ongsulee P. Artificial intelligence, machine learning and deep learning. 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE) 2017:1-6.
25. Pereira J, Silveira M. Learning Representations from Healthcare Time Series Data for Unsupervised Anomaly Detection. 2019 IEEE International Conference on Big Data and Smart Computing (BigComp)2019:1-7.
26. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty* 2018;33(8):2358-2361. (In eng). DOI: 10.1016/j.arth.2018.02.067.
27. Mazurowski MA, Buda M, Saha A, Bashir MR. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *J Magn Reson Imaging* 2019;49(4):939-954. (In eng). DOI: 10.1002/jmri.26534.
28. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *Radiographics* 2017;37(2):505-515. (In eng). DOI: 10.1148/rg.2017160130.
29. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural. *Advances in Neural Information Processing Systems* 2014:3320-3328.
30. Razavian AS, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: an astounding baseline for recognition. *CoRR abs/1403.6382* (2014). arXiv preprint arXiv:14036382 2014.
31. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88. (In eng). DOI: 10.1016/j.media.2017.07.005.
32. Brett A, Miller CG, Hayes CW, et al. Development of a clinical workflow tool to enhance the detection of vertebral fractures: accuracy and precision evaluation. *Spine (Phila Pa 1976)* 2009;34(22):2437-43. (In eng). DOI: 10.1097/BRS.0b013e3181b2eb69.
33. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. *Journal of Medical Imaging and Radiation Sciences* 2019;50(4):477-487. DOI: 10.1016/j.jmir.2019.09.005.
34. Ruder S. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:160904747 2016.
35. Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol* 2018;73(5):439-445. (In eng). DOI: 10.1016/j.crad.2017.11.015.
36. Jacques T, Cardot N, Ventre J, Demondion X, Cotten A. Commercially-available AI algorithm improves radiologists' sensitivity for wrist and hand fracture detection on

- X-ray, compared to a CT-based ground truth. *European Radiology* 2024;34(5):2885-2894. DOI: 10.1007/s00330-023-10380-1.
37. Rajpurkar P, Irvin J, Bagul A, et al. Mura: Large dataset for abnormality detection in musculoskeletal radiographs. *arXiv preprint arXiv:171206957* 2017.
 38. Langerhuizen DWG, Bulstra AEJ, Janssen SJ, et al. Is Deep Learning On Par with Human Observers for Detection of Radiographically Visible and Occult Fractures of the Scaphoid? *Clin Orthop Relat Res* 2020;478(11):2653-2659. (In eng). DOI: 10.1097/corr.0000000000001318.
 39. Dao T, Gu A, Ratner AJ, Smith V, De Sa C, Ré C. A Kernel Theory of Modern Data Augmentation. *Proc Mach Learn Res* 2019;97:1528-1537. (In eng).
 40. Yoon AP, Lee YL, Kane RL, Kuo CF, Lin C, Chung KC. Development and Validation of a Deep Learning Model Using Convolutional Neural Networks to Identify Scaphoid Fractures in Radiographs. *JAMA Netw Open* 2021;4(5):e216096. (In eng). DOI: 10.1001/jamanetworkopen.2021.6096.
 41. Beyaz S, Açıcı K, Sümer E. Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches. *Jt Dis Relat Surg* 2020;31(2):175-183. (In eng). DOI: 10.5606/ehc.2020.72163.
 42. Olczak J, Fahlberg N, Maki A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop* 2017;88(6):581-586. (In eng). DOI: 10.1080/17453674.2017.1344459.
 43. Lee K-C, Choi IC, Kang CH, et al. Clinical Validation of an Artificial Intelligence Model for Detecting Distal Radius, Ulnar Styloid, and Scaphoid Fractures on Conventional Wrist Radiographs. *Diagnostics* 2023;13(9):1657. (<https://www.mdpi.com/2075-4418/13/9/1657>
<https://pmc.ncbi.nlm.nih.gov/articles/PMC10178713/pdf/diagnostics-13-01657.pdf>).
 44. Kitamura G, Chung CY, Moore BE, 2nd. Ankle Fracture Detection Utilizing a Convolutional Neural Network Ensemble Implemented with a Small Sample, De Novo Training, and Multiview Incorporation. *J Digit Imaging* 2019;32(4):672-677. (In eng). DOI: 10.1007/s10278-018-0167-7.
 45. Lee NJ, Kothari P, Phan K, et al. Incidence and risk factors for 30-day unplanned readmissions after elective posterior lumbar fusion. *Spine* 2018;43(1):41-48.
 46. Yamamoto A, Fujita K, Yamada E, et al. Gait characteristics in patients with distal radius fracture using an in-shoe inertial measurement system at various gait speeds. *Gait & Posture* 2024;107:317-323.
 47. Kamer L, Noser H, Arand C, Handrich K, Rommens PM, Wagner D. Artificial intelligence and CT-based 3D statistical modeling to assess transsacral corridors and plan implant positioning. *Journal of Orthopaedic Research®* 2021;39(12):2681-2692.
 48. Poduval M, Ghose A, Manchanda S, Bagaria V, Sinha A. Artificial intelligence and machine learning: a new disruptive force in orthopaedics. *Indian journal of orthopaedics* 2020;54(2):109-122.
 49. Black DM, Rosen CJ. Clinical Practice. Postmenopausal Osteoporosis. *N Engl J Med* 2016;374(3):254-62. (In eng). DOI: 10.1056/NEJMc1513724.
 50. Pinto A, Reginelli A, Pinto F, et al. Errors in imaging patients in the emergency setting. *Br J Radiol* 2016;89(1061):20150914. (In eng). DOI: 10.1259/bjr.20150914.
 51. Warin K, Limprasert W, Suebnukarn S, Inglam S, Jantana P, Vicharueang S. Assessment of deep convolutional neural network models for mandibular fracture detection in panoramic radiographs. *International Journal of Oral and Maxillofacial Surgery* 2022;51(11):1488-1494.
 52. Murata K, Endo K, Aihara T, et al. Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. *Scientific Reports* 2020;10(1):20031.
 53. Inagaki N, Nakata N, Ichimori S, Udaoka J, Mandai A, Saito M. Detection of sacral fractures on radiographs using artificial intelligence. *JBJS Open Access* 2022;7(3):e22.

54. Wu J, Liu N, Li X, et al. Convolutional neural network for detecting rib fractures on chest radiographs: a feasibility study. *BMC Medical Imaging* 2023;23(1):18.
55. Choi JW, Cho YJ, Lee S, et al. Using a dual-input convolutional neural network for automated detection of pediatric supracondylar fracture on conventional radiography. *Investigative radiology* 2020;55(2):101-110.
56. Liu P-r, Zhang J-y, Xue M-d, et al. Artificial intelligence to diagnose tibial plateau fractures: an intelligent assistant for orthopedic physicians. *Current Medical Science* 2021;41(6):1158-1164.
57. Ashkani-Esfahani S, Yazdi RM, Bhimani R, et al. Detection of ankle fractures using deep learning algorithms. *Foot and Ankle Surgery* 2022;28(8):1259-1265.
58. Prijs J, Liao Z, To M-S, et al. Development and external validation of automated detection, classification, and localization of ankle fractures: inside the black box of a convolutional neural network (CNN). *European journal of trauma and emergency surgery* 2023;49(2):1057-1069.
59. Rundgren J, Bojan A, Mellstrand Navarro C, Enocson A. Epidemiology, classification, treatment and mortality of distal radius fractures in adults: an observational study of 23,394 fractures from the national Swedish fracture register. *BMC Musculoskelet Disord* 2020;21(1):88. (In eng). DOI: 10.1186/s12891-020-3097-8.
60. Crowe CS, Massenburg BB, Morrison SD, et al. Global trends of hand and wrist trauma: a systematic analysis of fracture and digit amputation using the Global Burden of Disease 2017 Study. *Inj Prev* 2020;26(Suppl 1):i115-i124. (In eng). DOI: 10.1136/injuryprev-2019-043495.
61. Nellans KW, Kowalski E, Chung KC. The epidemiology of distal radius fractures. *Hand Clin* 2012;28(2):113-25. (In eng). DOI: 10.1016/j.hcl.2012.02.001.
62. Rhemrev SJ, Ootes D, Beeres FJ, Meylaerts SA, Schipper IB. Current methods of diagnosis and treatment of scaphoid fractures. *Int J Emerg Med* 2011;4:4. (In eng). DOI: 10.1186/1865-1380-4-4.
63. Levin LS, Rozell JC, Pulos N. Distal Radius Fractures in the Elderly. *J Am Acad Orthop Surg* 2017;25(3):179-187. (In eng). DOI: 10.5435/jaaos-d-15-00676.
64. Liporace FA, Adams MR, Capo JT, Koval KJ. Distal radius fractures. *J Orthop Trauma* 2009;23(10):739-48. (In eng). DOI: 10.1097/BOT.0b013e3181ba46d3.
65. Prommersberger KJ, Fernandez DL. Nonunion of distal radius fractures. *Clin Orthop Relat Res* 2004(419):51-6. (In eng). DOI: 10.1097/00003086-200402000-00009.
66. Min H, Rabi Y, Wadhawan A, et al. Automatic classification of distal radius fracture using a two-stage ensemble deep learning framework. *Phys Eng Sci Med* 2023;46(2):877-886. (In eng). DOI: 10.1007/s13246-023-01261-4.
67. Anttila TT, Karjalainen TV, Mäkelä TO, et al. Detecting Distal Radius Fractures Using a Segmentation-Based Deep Learning Model. *J Digit Imaging* 2023;36(2):679-687. (In eng). DOI: 10.1007/s10278-022-00741-5.
68. Liu X, Gao K, Liu B, et al. Advances in Deep Learning-Based Medical Image Analysis. *Health Data Sci* 2021;2021:8786793. (In eng). DOI: 10.34133/2021/8786793.
69. Erdaş Ç B. Automated fracture detection in the ulna and radius using deep learning on upper extremity radiographs. *Jt Dis Relat Surg* 2023;34(3):598-604. (In eng). DOI: 10.52312/jdrs.2023.1312.
70. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep* 2017;7(1):5467. (In eng). DOI: 10.1038/s41598-017-05848-2.
71. Reigstad O, Grimsgaard C, Thorkildsen R, Reigstad A, Røkkum M. SCAPHOID NON-UNIONS, WHERE DO THEY COME FROM? THE EPIDEMIOLOGY AND INITIAL PRESENTATION OF 270 SCAPHOID NON-UNIONS. *Hand Surgery* 2012;17(03):331-335. DOI: 10.1142/s0218810412500268.
72. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *International conference on machine learning: PMLR*; 2019:6105-6114.

73. Gan K, Xu D, Lin Y, et al. Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. *Acta Orthop* 2019;90(4):394-400. (In eng). DOI: 10.1080/17453674.2019.1600125.
 74. Suzuki T, Maki S, Yamazaki T, et al. Detecting Distal Radial Fractures from Wrist Radiographs Using a Deep Convolutional Neural Network with an Accuracy Comparable to Hand Orthopedic Surgeons. *J Digit Imaging* 2022;35(1):39-46. (In eng). DOI: 10.1007/s10278-021-00519-1.
 75. Blüthgen C, Becker AS, Vittoria de Martini I, Meier A, Martini K, Frauenfelder T. Detection and localization of distal radius fractures: Deep learning system versus radiologists. *Eur J Radiol* 2020;126:108925. (In eng). DOI: 10.1016/j.ejrad.2020.108925.
 76. Zhang J, Li Z, Lin H, et al. Deep learning assisted diagnosis system: improving the diagnostic accuracy of distal radius fractures. *Front Med (Lausanne)* 2023;10:1224489. (In eng). DOI: 10.3389/fmed.2023.1224489.
 77. Brauer CA, Coca-Perrillon M, Cutler DM, Rosen AB. Incidence and mortality of hip fractures in the United States. *Jama* 2009;302(14):1573-1579.
 78. Center JR, Nguyen TV, Schneider D, Sambrook PN, Eisman JA. Mortality after all major types of osteoporotic fracture in men and women: an observational study. *The Lancet* 1999;353(9156):878-882.
 79. Ben-Shlomo Y, Blom A, Boulton C, et al. National Joint Registry Annual Reports. The National Joint Registry 17th Annual Report 2020. London: National Joint Registry
- © National Joint Registry 2020.; 2020.
80. Abdel MP, Watts CD, Houdek MT, Lewallen DG, Berry DJ. Epidemiology of periprosthetic fracture of the femur in 32 644 primary total hip arthroplasties. *The Bone & Joint Journal* 2016;98-B(4):461-467. DOI: 10.1302/0301-620x.98b4.37201.
 81. Ramavath A, Lamb JN, Palan J, Pandit HG, Jain S. Postoperative periprosthetic femoral fracture around total hip replacements: current concepts and clinical outcomes. *EFORT Open Rev* 2020;5(9):558-567. (In eng). DOI: 10.1302/2058-5241.5.200003.
 82. Capone A, Congia S, Civinini R, Marongiu G. Periprosthetic fractures: epidemiology and current treatment. *Clin Cases Miner Bone Metab* 2017;14(2):189-196. (In eng). DOI: 10.11138/ccmbm/2017.14.1.189.
 83. Zdolsek G, Chen Y, Bögl HP, Wang C, Woisetschlager M, Schilcher J. Deep neural networks with promising diagnostic accuracy for the classification of atypical femoral fractures. *Acta Orthop* 2021;92(4):394-400. (In eng). DOI: 10.1080/17453674.2021.1891512.
 84. Bégin MJ, Audet MC, Chevalley T, et al. Fracture Risk Following an Atypical Femoral Fracture. *J Bone Miner Res* 2022;37(1):87-94. (In eng). DOI: 10.1002/jbmr.4461.
 85. Cheung AM, McKenna MJ, van de Laarschot DM, et al. Detection of Atypical Femur Fractures. *J Clin Densitom* 2019;22(4):506-516. (In eng). DOI: 10.1016/j.jocd.2019.07.003.
 86. Kim T, Moon NH, Goh TS, Jung ID. Detection of incomplete atypical femoral fracture on anteroposterior radiographs via explainable artificial intelligence. *Sci Rep* 2023;13(1):10415. (In eng). DOI: 10.1038/s41598-023-37560-9.
 87. Alzaid A, Wignall A, Dogramadzi S, Pandit H, Xie SQ. Automatic detection and classification of peri-prosthetic femur fracture. *Int J Comput Assist Radiol Surg* 2022;17(4):649-660. (In eng). DOI: 10.1007/s11548-021-02552-5.
 88. Miao Y, Zhao P-F, Tang X-F, et al. A Method for Detecting Femur Fracture Based on SK-DenseNet. *Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing*. Dublin, Ireland: Association for Computing Machinery; 2019:Article 71.

89. Ouyang CH, Chen CC, Tee YS, et al. The Application of Design Thinking in Developing a Deep Learning Algorithm for Hip Fracture Detection. *Bioengineering (Basel)* 2023;10(6) (In eng). DOI: 10.3390/bioengineering10060735.
90. Urakawa T, Tanaka Y, Goto S, Matsuzawa H, Watanabe K, Endo N. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. *Skeletal Radiol* 2019;48(2):239-244. (In eng). DOI: 10.1007/s00256-018-3016-3.
91. Cheng CT, Wang Y, Chen HW, et al. A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs. *Nat Commun* 2021;12(1):1066. (In eng). DOI: 10.1038/s41467-021-21311-3.
92. Krogue JD, Cheng KV, Hwang KM, et al. Automatic Hip Fracture Identification and Functional Subclassification with Deep Learning. *Radiol Artif Intell* 2020;2(2):e190023. (In eng). DOI: 10.1148/ryai.2020190023.
93. Dominguez S, Liu P, Roberts C, Mandell M, Richman PB. Prevalence of traumatic hip and pelvic fractures in patients with suspected hip fracture and negative initial standard radiographs--a study of emergency department patients. *Acad Emerg Med* 2005;12(4):366-9. (In eng). DOI: 10.1197/j.aem.2004.10.024.
94. Mawatari T, Hayashida Y, Katsuragawa S, et al. The effect of deep convolutional neural networks on radiologists' performance in the detection of hip fractures on digital pelvic radiographs. *Eur J Radiol* 2020;130:109188. (In eng). DOI: 10.1016/j.ejrad.2020.109188.
95. Thian Y, Li Y, Jagmohan P, Sia D, Chan V, Tan R. Convolutional Neural Networks for Automated Fracture Detection and Localization on Wrist Radiographs. *Radiology: Artificial Intelligence* 2019;1:e180001. DOI: 10.1148/ryai.2019180001.
96. Jiménez-Sánchez A, Kazi A, Albarqouni S, et al. Precise proximal femur fracture classification for interactive training and surgical planning. *Int J Comput Assist Radiol Surg* 2020;15(5):847-857. (In eng). DOI: 10.1007/s11548-020-02150-x.
97. Tanzi L, Vezzetti E, Moreno R, Aprato A, Audisio A, Massè A. Hierarchical fracture classification of proximal femur X-Ray images using a multistage Deep Learning approach. *Eur J Radiol* 2020;133:109373. (In eng). DOI: 10.1016/j.ejrad.2020.109373.
98. Naranje SM, Erali RA, Warner WCJ, Sawyer JR, Kelly DM. Epidemiology of Pediatric Fractures Presenting to Emergency Departments in the United States. *Journal of Pediatric Orthopaedics* 2016;36(4):e45-e48. DOI: 10.1097/bpo.0000000000000595.
99. Nguyen T, Maarek R, Hermann AL, et al. Assessment of an artificial intelligence aid for the detection of appendicular skeletal fractures in children and young adults by senior and junior radiologists. *Pediatr Radiol* 2022;52(11):2215-2226. (In eng). DOI: 10.1007/s00247-022-05496-3.
100. Praveen MY, Adriel Guang Wei G, Sze Ying Y, et al. Accuracy of radiologists and radiology residents in detection of paediatric appendicular fractures with and without artificial intelligence. *BMJ Health & Care Informatics* 2024;31(1):e101091. DOI: 10.1136/bmjhci-2024-101091.
101. George MP, Bixby S. Frequently Missed Fractures in Pediatric Trauma: A Pictorial Review of Plain Film Radiography. *Radiol Clin North Am* 2019;57(4):843-855. (In eng). DOI: 10.1016/j.rcl.2019.02.009.
102. Zech JR, Carotenuto G, Igbinoba Z, et al. Detecting pediatric wrist fractures using deep-learning-based object detection. *Pediatric Radiology* 2023;53(6):1125-1134. DOI: 10.1007/s00247-023-05588-8.
103. Zech JR, Ezuma CO, Patel S, et al. Artificial intelligence improves resident detection of pediatric and young adult upper extremity fractures. *Skeletal Radiology* 2024;53(12):2643-2651. DOI: 10.1007/s00256-024-04698-0.
104. Aryasomayajula S, Hing C, Siebachmeyer M, et al. Developing an artificial intelligence diagnostic tool for paediatric distal radius fractures, a proof of concept

- study. *The Annals of The Royal College of Surgeons of England* 2023;105(8):721-728. DOI: 10.1308/rcsann.2023.0017 %M 37642151.
105. Ghosh A, Bose S, Patton D, et al. Deep learning-based prediction of rib fracture presence in frontal radiographs of children under two years of age: a proof-of-concept study. *British Journal of Radiology* 2023;96(1145). DOI: 10.1259/bjr.20220778.
 106. Nagy E, Janisch M, Hrzić F, Sorantin E, Tschauer S. A pediatric wrist trauma X-ray dataset (GRAZPEDWRI-DX) for machine learning. *Scientific Data* 2022;9(1):222. DOI: 10.1038/s41597-022-01328-z.
 107. Binh LN, Nhu NT, Vy VPT, et al. Multi-Class Deep Learning Model for Detecting Pediatric Distal Forearm Fractures Based on the AO/OTA Classification. *Journal of Imaging Informatics in Medicine* 2024;37(2):725-733. DOI: 10.1007/s10278-024-00968-4.
 108. Knight J, Zhou Y, Keen C, et al. 2D/3D ultrasound diagnosis of pediatric distal radius fractures by human readers vs artificial intelligence. *Scientific Reports* 2023;13(1):14535. DOI: 10.1038/s41598-023-41807-w.
 109. Medical HY. DR AI-assisted Bone Fracture Detection. 2023 (<https://radiology.healthairegister.com/products/hy-medical-dr-bone-fracture/>).
 110. Duron L, Ducarouge A, Gillibert A, et al. Assessment of an AI Aid in Detection of Adult Appendicular Skeletal Fractures by Emergency Physicians and Radiologists: A Multicenter Cross-sectional Diagnostic Study. *Radiology* 2021;300(1):120-129. (In eng). DOI: 10.1148/radiol.2021203886.
 111. Meetschen M, Salhöfer L, Beck N, et al. AI-Assisted X-ray Fracture Detection in Residency Training: Evaluation in Pediatric and Adult Trauma Patients. *Diagnostics (Basel)* 2024;14(6) (In eng). DOI: 10.3390/diagnostics14060596.
 112. Dell'Aria A, Tack D, Saddiki N, et al. Radiographic Detection of Post-Traumatic Bone Fractures: Contribution of Artificial Intelligence Software to the Analysis of Senior and Junior Radiologists. *J Belg Soc Radiol* 2024;108(1):44. (In eng). DOI: 10.5334/jbsr.3574.
 113. Cohen M, Puntinet J, Sanchez J, et al. Artificial intelligence vs. radiologist: accuracy of wrist fracture detection on radiographs. *Eur Radiol* 2023;33(6):3974-3983. (In eng). DOI: 10.1007/s00330-022-09349-3.
 114. Oppenheimer J, Lüken S, Hamm B, Niehues SM. A Prospective Approach to Integration of AI Fracture Detection Software in Radiographs into Clinical Workflow. *Life (Basel)* 2023;13(1) (In eng). DOI: 10.3390/life13010223.
 115. Guermazi A, Tannoury C, Kompel AJ, et al. Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence. *Radiology* 2022;302(3):627-636. (In eng). DOI: 10.1148/radiol.210937.
 116. Hayashi D, Kompel AJ, Ventre J, et al. Automated detection of acute appendicular skeletal fractures in pediatric patients using deep learning. *Skeletal Radiol* 2022;51(11):2129-2139. (In eng). DOI: 10.1007/s00256-022-04070-0.
 117. Altmann-Schneider I, Kellenberger CJ, Pistorius S-M, et al. Artificial intelligence-based detection of paediatric appendicular skeletal fractures: performance and limitations for common fracture types and locations. *Pediatric Radiology* 2024;54(1):136-145. DOI: 10.1007/s00247-023-05822-3.
 118. Schweikhard FP, Kosanke A, Lange S, et al. Doctor's Orders—Why Radiologists Should Consider Adjusting Commercial Machine Learning Applications in Chest Radiography to Fit Their Specific Needs. *Healthcare* 2024;12(7):706. (<https://www.mdpi.com/2227-9032/12/7/706>).
 119. Gipson J, Tang V, Seah J, et al. Diagnostic accuracy of a commercially available deep-learning algorithm in supine chest radiographs following trauma. *British Journal of Radiology* 2022;95(1134). DOI: 10.1259/bjr.20210979.
 120. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* 2018;392(10162):2388-2396. (In eng). DOI: 10.1016/s0140-6736(18)31645-3.

121. Buchlak QD, Tang CHM, Seah JCY, et al. Effects of a comprehensive brain computed tomography deep learning model on radiologist detection accuracy. *European Radiology* 2024;34(2):810-822. DOI: 10.1007/s00330-023-10074-8.
122. Dupuis M, Delbos L, Rouquette A, Adamsbaum C, Veil R. External validation of an artificial intelligence solution for the detection of elbow fractures and joint effusions in children. *Diagnostic and Interventional Imaging* 2024;105(3):104-109. DOI: <https://doi.org/10.1016/j.diii.2023.09.008>.
123. Adams M, Chen W, Holcdorf D, McCusker MW, Howe PD, Gaillard F. Computer vs human: Deep learning versus perceptual training for the detection of neck of femur fractures. *J Med Imaging Radiat Oncol* 2019;63(1):27-32. (In eng). DOI: 10.1111/1754-9485.12828.
124. Willemink MJ, Koszek WA, Hardell C, et al. Preparing Medical Imaging Data for Machine Learning. *Radiology* 2020;295(1):4-15. (In eng). DOI: 10.1148/radiol.2020192224.
125. Cheng CT, Ho TY, Lee TY, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. *Eur Radiol* 2019;29(10):5469-5477. (In eng). DOI: 10.1007/s00330-019-06167-y.
126. Parliament E. Lei da UE sobre IA: primeira regulamentação de inteligência artificial. European Parliament. (<https://www.europarl.europa.eu/topics/pt/article/20230601STO93804/lei-da-ue-sobre-ia-primeira-regulamentacao-de-inteligencia-artificial>)- consultado pela última vez em dezembro/2024
127. Banatwala U, Ibrahim MT, Shaikh RH, Shahzad H, Hoodbhoy Z, Noordin S. A comprehensive exploration of artificial intelligence in orthopaedics within lower-middle-income countries: a narrative review. *J Pak Med Assoc* 2024;74(4 (Supple-4)):S90-s96. (In eng). DOI: 10.47391/jpma.Aku-9s-14.

