

Algoritmo de Aprendizagem Automática na Classificação de Radiografia ao Tórax em Incidência frontal como “normais” ou “patológicos”

Machine Learning Algorithm in the Classification of frontal view Chest x-ray as “normal” or “pathological”

Ivone Cidade de Moura¹, Luis Mesquita², Ricardo Teresa Ribeiro³

1 Técnica Superior de Radiologia - Centro Hospitalar Universitário Lisboa Norte, Instituto Superior de Engenharia de Lisboa

2 Técnico de Electromedicina – Multipore, Instituto Superior de Engenharia de Lisboa

3 H&TRC - Health & Technology Research Center, ESTeSL - Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa.

Resumo

A crescente procura de exames radiológicos pressiona os serviços de imagiologia que enfrentam desafios devido à escassez de radiologistas, exigindo uma interpretação mais rápida com um maior potencial de erro. Em paralelo, as técnicas de *radiomics* e inteligência artificial (IA) têm-se verificado ferramentas importantes no campo da radiologia, revolucionando a própria prática clínica. Através destas técnicas, este artigo desenvolveu um classificador de deteção de patologia para radiografias do tórax, da base de dados pública ChestXray14, com o objetivo de realçar o papel crucial que a compreensão das técnicas de *radiomics* e IA na profissão do técnico de radiologista.

Foram analisadas 1662 radiografias (50% com patologia) ao tórax, tendo sido aplicadas duas estratégias para seleção de 5 características *radiomics*: (i) análise de componentes principais (PCA) e (ii) “*information gain ratio*” (Rank), utilizando o software Orange. Com o método PCA, com redução para 5 componentes e 73% de variância explicada, o melhor classificador foi o *Neural Network*, com 0,987 de *Area Under the Curve* (AUC). No *Information Gain Ratio* também o *Neural Network* foi o melhor classificador com 0,972 AUC, na qual se verificou uma sensibilidade de 97,8%, especificidade de 92,9% e precisão de 93%.

Ao utilizar técnicas de IA e tirar partido de um grande conjunto de dados, o nosso estudo demonstra a viabilidade da utilização de classificadores automáticos para ajudar na interpretação de radiografias ao tórax, indicando o seu potencial como uma ferramenta valiosa na triagem, priorização de exames, e otimização no fluxo de trabalho nos departamentos de radiologia.

Abstract

The growing demand for radiological exams puts pressure on imaging services that face challenges due to the shortage of radiologists, requiring faster interpretation with a higher potential for error. In parallel, radiomics and artificial intelligence (AI) techniques have proven to be important tools in the field of radiology, revolutionizing clinical practice itself. Using these techniques, this paper developed a pathology detection classifier for chest radiographs, from the public database ChestXray14, in order to highlight the crucial role that understanding radiomics techniques and AI play in the radiographer's profession.

We analysed 1662 chest radiographs (50% with pathology) and applied two strategies for selecting 5 radiomics features: (i) principal component analysis and (ii) information gain ratio, using the Orange software. With the PCA method, with reduction to 5 components and 73% of variance explained, the best classifier was the Neural Network, with 0.987 of Area Under the Curve (AUC). In Information Gain Ratio also the Neural Network was the best classifier with 0.972 AUC, in which a sensitivity of 97.8%, specificity of 92.9% and accuracy of 93% were found.

By using AI techniques and taking advantage of a large dataset, our study demonstrates the feasibility of using automatic classifiers to aid in the interpretation of chest radiographs, indicating their potential as a valuable tool in screening, prioritizing exams, and optimizing workflow in radiology departments.

Palavras-chave: inteligência artificial, *computer-aided detection*, classificadores de aprendizagem automática, radiografia tórax, *principal component analysis*, rank

Keywords: artificial intelligence, computer-aided detection, machine learning classifiers, chest x-ray, principal component analysis, rank

Introdução

Os avanços tecnológicos aumentaram significativamente a acessibilidade e a capacidade de produção de exames radiológicos, reduzindo simultaneamente os tempos de exame.¹ A transição digital dos serviços de radiologia facilitou ainda mais o arquivo, o processamento e extração de informações de cariz clínico dos exames radiológicos, permitindo o acompanhamento e análise longitudinal dos doentes.² Contudo, o crescente volume de exames ultrapassou a disponibilidade dos serviços, levando à redução do tempo de interpretação e ao aumento da pressão para fornecer pareceres clínicos atempados, particularmente em ambiente de urgência médica.^{3,4}

Com o objetivo de otimizar a alocação de recursos, os radiologistas geralmente priorizam os exames que exigem interpretação mais especializada e contam com modalidades de imagem mais simples, como a radiografia, para serem lidas por médicos sem ampla experiência radiológica.^{3,4} Esta abordagem ajuda a mitigar o risco de má interpretação e garante relatórios precisos em exames mais complexos, salvaguardando, em última análise, os resultados dos pacientes. De facto, o processo e resultado da análise dos exames radiológicos está dependente da experiência, formação e níveis de atenção dos radiologistas. Dado o elevado volume de exames diários, é possível que padrões anormais subtis passem despercebidos. Assim, a integração de métodos de inteligência artificial (IA) na prática clínica tem o potencial de aumentar a eficiência da deteção de anomalias e reduzir o tempo de resposta.^{3,5}

A IA surge como uma das principais plataformas transformadoras nos cuidados de saúde, mantendo como estratégia apoiar a resolução de diferentes desafios clínicos através da utilização de dados clínicos (ex. radiografia tórax) e algoritmos de classificação. Para garantir um impacto positivo, deve-se considerar uma estrutura de cuidados de saúde baseada na criação de valor, assumindo-se como valor o rácio entre os resultados clínicos e os seus custos associados. Neste contexto, a IA pode gerar valor reduzindo os custos e/ou melhorando os resultados em termos de saúde.⁶ Na radiologia, a IA pode ser utilizada para melhorar a eficiência do fluxo de trabalho, acelerar o tempo de leitura do exame, minimizar a dose de radiação e a utilização de agentes de contraste, melhorar a precisão do diagnóstico para a deteção precoce de doenças e permitir diagnósticos personalizados.⁷⁻⁹

Diversos estudos têm demonstrado aplicações bem-sucedidas de IA na classificação de radiografias ao tórax, na qual se salienta o seu potencial para melhorar a precisão do diagnóstico e ajudar os profissionais de saúde no seu processo de tomada de decisões. Por exemplo, Wang *et al.* (2017) desenvolveram um modelo de aprendizagem profunda denominado ChestX-ray8 para classificar automaticamente doenças comuns do tórax. O seu modelo obteve resultados promissores com uma *Area Under the Curve* (AUC) multirrótulo de 0,809 e superou as abordagens tradicionais de aprendizagem automática. Rajpurkar *et al.* (2017) introduziram a CheXNet, uma rede neural convolucional (CNN), que demonstrou uma elevada precisão na identificação de 14 patologias diferentes em radiografias do tórax, superando o desempenho dos radiologistas. Do mesmo modo, Yao *et al.* (2020) utilizaram um modelo de aprendizagem profunda denominado COVID-Net para detetar casos de COVID-19 em radiografias de tórax. A sua abordagem alcançou uma sensibilidade de 94,1% e uma especificidade de 92,7% na deteção da COVID-19, mostrando potencial para ajudar no diagnóstico desta doença específica. Num estudo recente, Dyer *et al.*, demonstrou que um algoritmo de Deep Learning foi capaz de classificar 15% de radiografias ao tórax como “normais” com uma precisão de 97,7% e uma taxa de erro de 0,33%, removendo 24,9% dos estudos normais confirmados pelo radiologista do fluxo de trabalho.^{5,10}

Mais recentemente outra técnica – *radiomics* – oriunda da área de *computer vision* da engenharia tem vindo a ser aplicada com sucesso na área da radiologia. *Radiomics* refere-se à extração e análise de uma vasta gama de características quantitativas de imagens médicas.⁶ Permite a conversão de imagens radiográficas convencionais em

dados de diferentes dimensões e reprodutíveis, fornecendo informações correlacionáveis com o diagnóstico de doenças, o prognóstico e a resposta ao tratamento. Ao compreender a mais-valia da extração de características *radiômicas*, os técnicos de radiologia podem tirar partido do poder destas características correlacionando os padrões técnicos e semiológicos de análise das imagens radiográficas, contribuindo assim para melhorar os cuidados prestados aos doentes e a tomada de decisões clínicas.¹¹

Com base nos desenvolvimentos observados, a compreensão da técnica de extração de características radiômicas e a utilização de IA em imagens médicas é de extrema importância para os técnicos de radiologia, não só para a melhoria contínua das suas funções enquanto profissionais de saúde, mas também na capacidade de um desenvolvimento mais centrado no doente, bem como a integração em equipas de investigação, tornando-se assim como colaboradores indispensáveis na área em constante evolução da imagem médica. Este artigo tem assim como objetivos (i) desenvolver um classificador de deteção de patologia para a radiografias do tórax, da base-de-dados pública CestXray14, e (ii), com base nos resultados, realçar o papel crucial que a compreensão das técnicas de *radiomics* e IA podem ter na profissão do técnico de radiologia.

Materiais e Métodos

A metodologia de classificação supervisionada através da extração de características radiômicas de radiografias do tórax proposta neste artigo tem os seguintes passos principais:

1. Base de dados: Um conjunto de radiografias do tórax é extraído e organizado num formato adequado para análise e reprodutibilidade dos resultados.
2. Segmentação da região de interesse: A região de interesse (ROI), que neste caso é o parênquima pulmonar, é segmentada a partir das imagens radiográficas.
3. Extração de características *radiomics*: Diversas características radiômicas são extraídas da região de interesse segmentada.
4. Seleção de características: Com base nas características extraídas, um processo de seleção é realizado para identificar as características mais relevantes para a classificação das radiografias do tórax.
5. Treino de diferentes modelos de classificação: diferentes algoritmos de classificação supervisionada são treinados utilizando as características seleccionadas e as classes ("patológico" e "normal") associadas a cada radiografia do tórax. O processo de treino envolve "alimentar" o modelo com diferentes partições de dados e ajustar seus parâmetros para otimizar a capacidade de prever corretamente as classes.
6. Avaliação da performance modelo: O modelo treinado é avaliado usando um conjunto de dados de teste separado. Isso permite verificar o desempenho do modelo em radiografias do tórax não vistas durante o treino. Métricas de desempenho, como AUC, sensibilidade e especificidade, são calculadas para avaliar a capacidade de classificação do modelo.

Base de dados

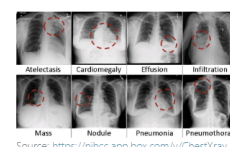
Foi utilizada a base-de-dados pública ChestX-ray14, obtida através do *website* [Kaggle](https://www.kaggle.com/datasets/nihcc/chestx-ray14), possuindo 112120 imagens de radiografia do tórax em incidência frontal de 30805 pacientes.¹²

ChestX-ray14

Introduced by Wang et al. in ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases

ChestX-ray14 is a medical imaging dataset which comprises 112,120 frontal-view X-ray images of 30,805 (collected from the year of 1992 to 2015) unique patients with the text-mined fourteen common disease labels, mined from the text radiological reports via NLP techniques. It expands on ChestX-ray8 by adding six additional thorax diseases: Edema, Emphysema, Fibrosis, Pleural Thickening and Hernia.

Source: <https://nihcc.app.box.com/v/ChestXray-NIHCC/file/220660789610>



Source: <https://nihcc.app.box.com/v/ChestXray-NIHCC/file/220660789610>

Figura 1 - Base de dados ChestX-ray14.

A ChestX-ray14 apresenta algumas limitações: (i) diferentes faixas etárias - contém exames de crianças, jovens e adultos; (ii) apesar de todas serem um exame frontal ao tórax, a aquisição varia entre em incidência antero-posterior (AP) como postero-anterior (PA), o que se traduz numa variação das dimensões da anatomia e na sua proporção na relação com as restantes estruturas.

Para ultrapassar estas limitações, as imagens seleccionadas foram analisadas por uma técnica de radiologia com mais de 15 anos de experiência, levando assim a uma forte redução da quantidade de imagens analisadas. Foram assim seleccionadas 2000 radiografias ao tórax em posição frontal, onde 984 estavam identificadas com a presença de pelo menos uma patologia.

Seleção das Radiografias ao Tórax

A ChestX-ray14 é disponibilizada com informação acerca de cada uma das radiografias, nomeadamente: número de identificação (ID), categorização da patologia(s) identificadas e coordenadas da ROI (bounding box). As patologias estão organizadas nas seguintes categorias: Atelectasia, Perfusão, Massa, Nódulo, Pneumonia e Cardiomegália. A categoria “Cardiomegália” foi excluída do estudo por duas razões: (i) ao tratar-se de um aumento do aumento da sombra cardíaca, calculado através do rácio cardiorádico, não é expectável que existam alterações no parênquima pulmonar; (ii) a avaliação da silhueta cardíaca pode não permitir um diagnóstico diferencial entre as causas do aumento do contorno cardiomediastínico com o aumento da silhueta cardíaca e, em última análise, não se trata de uma patologia pulmonar.¹³ Esta exclusão reduziu a quantidade de imagens patológicas a analisar para 839, das 984 imagens patológicas inicialmente seleccionadas.

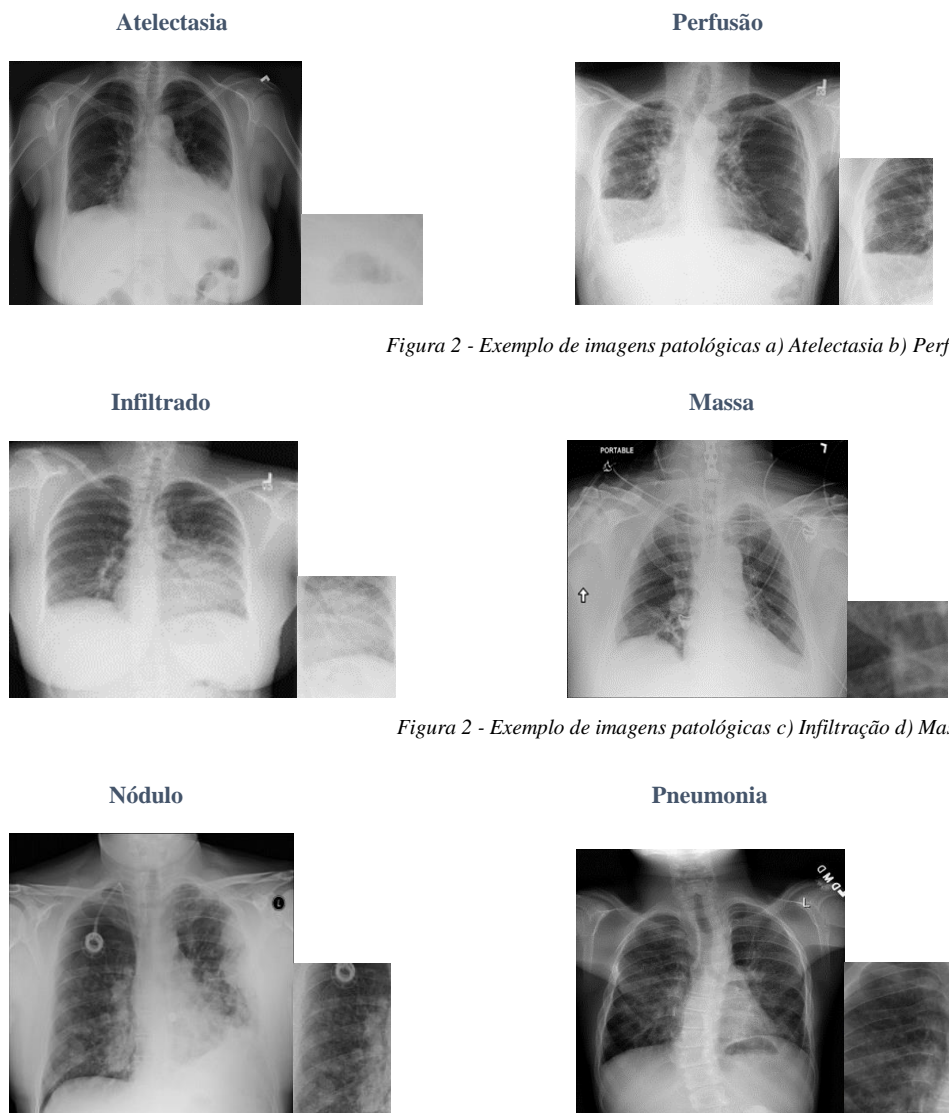


Figura 2 - Exemplo de imagens patológicas a) Atelectasia b) Perfusão.

Figura 2 - Exemplo de imagens patológicas c) Infiltração d) Massa.

Figura 2 - Exemplos de imagens patológicas e) Nódulo f) Pneumonia.

A seleção das imagens normais foi realizada através da aplicação de um algoritmo que extraiu aleatoriamente 900 imagens a partir do ficheiro que acompanha a BD e que estavam identificadas como normais. Procedeu-se a uma pré-análise na qual foram retiradas as imagens com algumas evidências patológicas ou alterações anatómicas muito significativas. Este processo reduziu as imagens patológicas para 813.

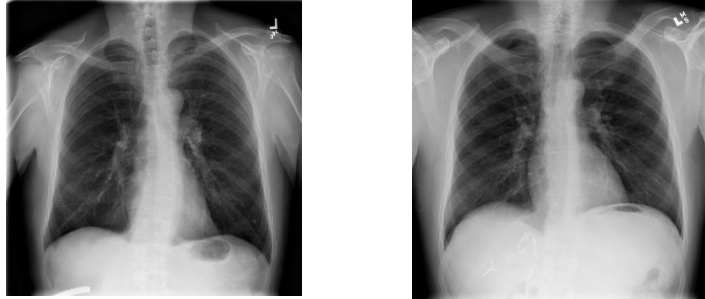


Figura 3 - Exemplos de imagens sem patologia.

Segmentação da Região de Interesse das Radiografias ao Tórax

As imagens patológicas e normais foram alvo de processamento. Em relação às imagens patológicas, iniciou-se pelo cálculo e determinação da máscara, que é necessário para a aplicação da função que irá extrair características. Para tal é necessário definir uma região de interesse, delimitada por uma caixa retangular que serve de ponto de referência para a deteção do objeto, denominada *bounding box*. A *bounding box* de cada imagem está identificada em coordenadas *xywh*. Para a criação da máscara binária, que vai segmentar a área do parênquima pulmonar que vai ser estudado, foram criados os limites *left*, *right*, *top* e *bottom*, tendo sido necessário converter a *bounding box* conforme descrito na figura 4.

$$\begin{aligned} \text{left} &= x, \quad \text{top} = y \\ \text{bottom} &= y - h, \quad \text{right} = x + w \end{aligned}$$

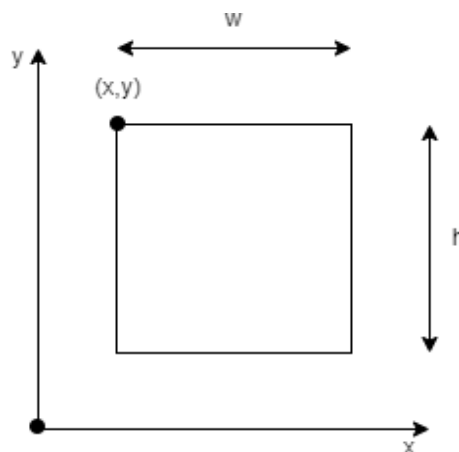


Figura 4 - Conversão de *xywh* para *left*, *right*, *top* e *bottom* para criação da *Bounding box* (ROI) que irá segmentar o parênquima pulmonar identificado com as diferentes patologias.

As imagens sem patologia não possuem *bounding box*. De forma a poder aplicar a extração de features criou-se uma máscara ou ROI genérica para ser aplicada em todas as 813 imagens normais. A ROI define a região do ápex pulmonar direito, por ser mais fácil de isolar da restante anatomia e permitir maior contraste entre estruturas anatómicas, diferenciando-se melhor das regiões mais suscetíveis a vestígios patológicos. A ROI aplicada às radiografias normais pode ser visualizada na figura seguinte.

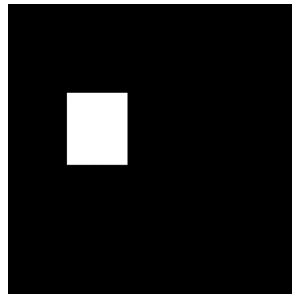


Figura 5 - ROI aplicada às imagens normais, na qual a região a branco será aquela que irá segmentar o parênquima pulmonar normal.

Extração de características radiomics das ROI selecionadas

Radiomics é um processo automático ou semiautomático de extração uma grande quantidade quantificável de características de uma região de interesse em imagens radiográficas. Este método atribui valores numéricos às características extraídas, sendo estas relacionadas com intensidade, textura, forma e análises em frequência. Assim podemos agrupar as referidas características em 3 grandes grupos: ¹⁴

1. Características de primeira ordem

Estas descrevem a distribuição da intensidade dos pixels da região de interesse, estando assim associadas ao histograma da imagem. Estão incluídas neste grupo a média, mediana e desvio padrão da intensidade dos pixels, bem como outras grandezas como energia (avalia a grandeza dos valores dos pixels), entropia (mede a incerteza na imagem, ou seja, quanto maior o valor, mais heterogéneo é o ROI), *kurtosis* (descreve a probabilidade de distribuição dos valores na ROI, ou seja, se se aproxima do valor médio ou não) e a *skewness* (mede a assimetria da probabilidade de distribuição do ROI em relação ao valor médio). ^{14,15}

2. Características morfológicas

As características de forma 2D e 3D englobam métodos descritivos da forma e volume da região de interesse, que são calculados a partir da imagem segmentada. Incluem o cálculo do volume, área, diâmetro máximo e outros parâmetros relacionados à geometria da ROI. A relação superfície-volume elevada indica uma área de estudo com geometria irregular ou espiculada, enquanto um valor mais baixo indica uma área de estudo com forma mais arredondada. ¹⁴

3. Características de textura

O termo "textura da imagem" refere-se à variação espacial dos níveis de cinza presentes na imagem. O conceito de textura intratumoral, que representa a heterogeneidade de uma área de interesse, foi introduzido por Halarick em 1973. ¹¹

As características de textura são calculadas por meio da aplicação de matrizes que avaliam as relações espaciais entre as intensidades de dois pixels, considerando uma determinada distância e direção. Essas matrizes permitem quantificar a homogeneidade ou heterogeneidade da distribuição dos níveis de cinza, fornecendo informações sobre a textura da imagem, tendo sido extraídas as seguintes: ¹⁴⁻¹⁶

- Gray Level Co-occurrence Matrix (GLCM) - Avalia a distribuição espacial dos níveis de cinzento, baseado na ocorrência repetida de alguns níveis de cinzento. Permite calcular 24 características.
- Gray Level Run Length Matrix (GLRLM) - avalia o número consecutivo de pixels com o mesmo nível de cinzento. Permite calcular 16 características.
- Gray Level Size Zone Matrix (GLSZM) permite avaliar o número de pixels consecutivos com o mesmo valor de nível de cinzento, ou seja, quantifica zonas homogêneas de cinzento.
- Gray Level Dependence Matrix (GLDM) avalia o número de pixels consecutivos dependentes do pixel central e quantifica o nível de cinzentos dependentes na ROI. Permite extrair 14 características.
- Neighbouring Gray Tone Difference Matrix (NGTDM) calcula a diferença do valor de cinzento entre um pixel e os seus vizinhos. Permite calcular 5 características.

Foram assim extraídas 123 características para cada uma das ROI selecionadas. Devido ao número elevado de características (ou *features*) e com o objetivo de evitar *overfitting*, foi necessário proceder a uma metodologia de redução do espaço de características.

Seleção das características

A etapa de seleção de características é crucial para evitar *overfitting* do modelo a ser desenvolvido, assim após a extração das características realizou-se uma pré-análise dos dados com o objetivo de reduzir a dimensionalidade do espaço de características: (i) análise de correlação - utilizando o coeficiente de correlação de Pearson optou-se por retirar uma das características que se relacionavam entre si com um coeficiente superior a 0.95 ($p \text{ value} < 0.01$); de seguida procedeu-se a uma (ii) análise de outliers – optou-se pelo método Local Outlier Factor (15% de contaminação, 20 vizinhos e métrica euclidiana) através da função outliers do programa Orange.

A partir das 46 características resultantes do processo de pré-análise, foram selecionadas as cinco características com maior importância, recorrendo a duas técnicas de seleção de características:

(i) Ranking estatístico information gain ratio. Este indicador considera tanto a relevância da característica em relação à variável de interesse como a quantidade de informações que ela fornece em relação a outras características. É calculado através da entropia dos dados calculada antes e depois de considerar cada uma das características, fornecendo uma medida da qualidade de informação ganha ao se adicionar uma determinada característica ao modelo.¹⁷ Foi utilizado a função “Rank” no programa Orange 3 (ver Fig. 6).

(ii) Análise por Componentes Principais (PCA) – Método que realiza diferentes rotações ao conjunto de características de forma a que as características rodadas não estejam estatisticamente correlacionadas. Esta rotação é seguida da seleção de um subconjunto de novas características, de acordo com a sua importância para a explicação dos dados.¹⁸ Neste projeto, optou-se por reduzir de 46 features para 5 componentes, com uma variância explicada de 73%. Foi utilizada a função “PCA” no programa Orange 3 (ver Fig. 6).

Processo de classificação e indicadores de performance

Para os diferentes conjuntos de características selecionados (Rank e PCA), e tratando-se de um processo de classificação supervisionado com duas classes, “normal”, “patológico”, foram testados diferentes classificadores: kNN, SVM, Random Forest e Neural Network.

O kNN é um método não-paramétrico que avalia a distância entre classes em relação aos k vizinhos mais próximos. O número de vizinhos utilizado foi de 5. O classificador *Random Forest* combina o resultado de múltiplas árvores de decisão, tendo como objetivo diminuir a variância entre os resultados através da criação de diferentes ligações entre as várias árvores de decisão com diversas ramificações. Neste projeto utilizaram-se 10 árvores de decisão. O classificador SVM (*Support Vector Machine*) é um robusto classificador baseado no princípio de *kernels*, que procura definir o hiper-plano que melhor separa os vetores de suporte de cada uma das classes consideradas.¹⁹ Foi utilizado o kernel RBF. O classificador Neural Network utilizado é um algoritmo *multi-layer-perception* (MLP) com *backpropagation* implementado no Orange 3. Foi parametrizado com um máximo de 200 iterações e 100 camadas de neurónios.

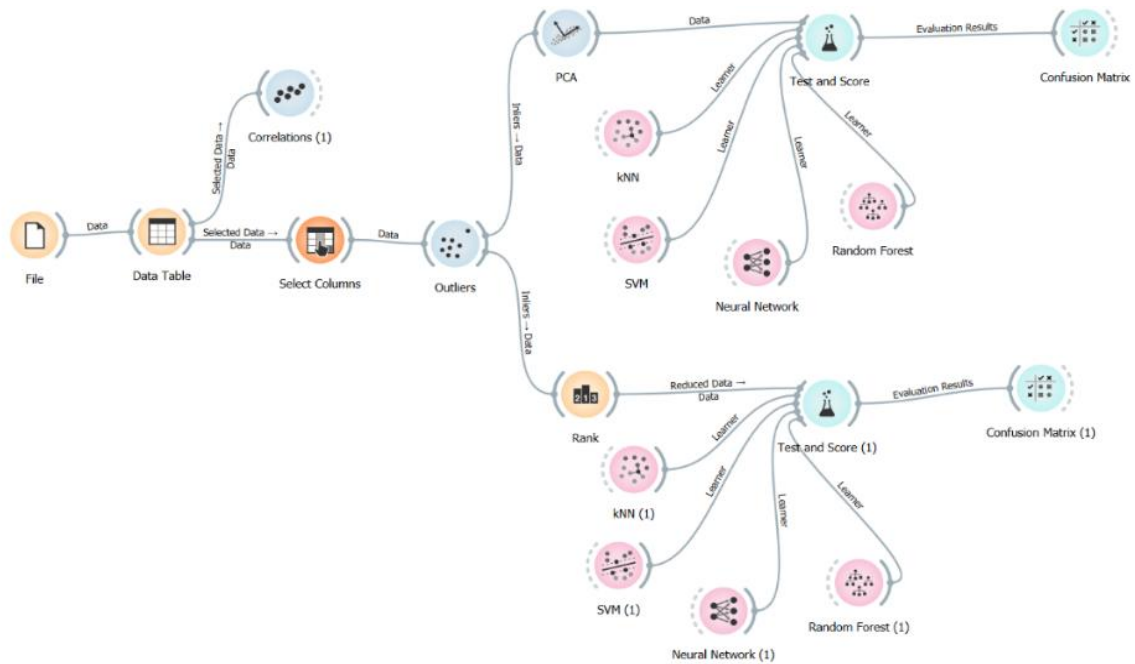


Figura 6 - Workflow de pre-análise e seleção de características, classificação e avaliação da performance dos diferentes classificadores, no programa Orange3.

Indicadores de performance utilizados

A AUC está associada à curva *Receiver Operating Characteristic* (ROC) e ambas são ferramentas de medição e comparação do desempenho de modelos de classificação binários. A curva ROC é uma ferramenta estatística em que as curvas são representadas num gráfico bidimensional, com a ordenada a representar a fração de verdadeiros positivos ou sensibilidade e a abcissa a representar a fração de falsos positivos ou 1-especificidade. A especificidade representa a proporção de casos negativos que foram corretamente identificados. A AUC é uma medida de área que facilita a comparação entre curvas ROC, sendo um indicador primário da performance dos classificadores. Quanto maior o AUC, melhor a performance do classificador.¹⁹

A avaliação da performance dos classificadores é realizada através da análise da matriz de confusão, nomeadamente de duas frações, relativas aos verdadeiros negativos (Fração de Verdadeiros Negativos, FVN) e aos falsos negativos (Fração de Falsos Negativos, FFN). No processo de classificação em diagnóstico médico, a FFN e a FFP são importantes devido a serem erros e consequentemente trazerem graves implicações para o indivíduo. Sob esta perspetiva, um falso negativo terá sempre mais impacto que um falso positivo, pois um “doente” ao ser inferido como um falso negativo não será sujeito a qualquer tipo de tratamento o que pode em casos extremos, levar à sua morte. Já um falso positivo não expressa consequências tão graves, pois a longo prazo o paciente terá menos sequelas que um falso negativo, por este necessitar de uma intervenção médica que não lhe foi fornecida.¹⁹ Após a definição dos métodos de PCA e Rank, avaliaram-se a performance de 4 modelos de classificação: kNN, SVM, Random Forest e Neural Network. Para a análise da matriz de confusão, centramo-nos em 3 parâmetros: a sensibilidade (taxa de verdadeiros positivos), a especificidade (taxa de verdadeiros negativos) e precisão (valor preditivo positivo). Os cálculos a efetuar estão exemplificados na figura seguinte:

	<i>Positivo</i>	<i>Negativo</i>	
<i>Positivo</i>	Verdadeiro Positivo (VP)	Falso Negativo (FN)	Sensibilidade $\frac{VP}{(VP + FN)}$
<i>Negativo</i>	Falso Positivo (FP)	Verdadeiro Negativo (VN)	Especificidade $\frac{VN}{(VN + FP)}$
	Precisão $\frac{VP}{(VP + FP)}$	Valor Preditivo Negativo $\frac{VN}{(VN + FN)}$	Exatidão $\frac{(VP + VN)}{(VP + VN + FP + FN)}$

Figura 7 - Cálculo da Precisão, Sensibilidade e Especificidade através da matrix de confusão.

Resultados e Discussão

Este artigo desenvolveu um classificador de detecção de patologia para radiografias do tórax, da base de dados pública ChestXray14, com o objetivo de realçar o papel crucial que a compreensão das técnicas de radiomics e IA na profissão do técnico de radiologista.

Foram analisadas 1662 radiografias (50% com patologia) ao tórax, tendo sido aplicadas duas estratégias para seleção de 5 características radiomics: (i) análise de componentes principais (PCA) e (ii) “information gain ratio” (Rank), utilizando o software Orange.

Os resultados obtidos pelo conjunto de características extraídas pelo método PCA encontram-se resumidos na tabela 1 e figura 8.

Tabela 1 – Resultados obtidos de AUC, Precisão, Sensibilidade e Especificidade em PCA

MODEL	AUC	PRECISÃO	SENSIBILIDADE	ESPECIFICIDADE
A) KNN	0,781	69,2 %	77,3 %	66,6 %
B) RANDOM FOREST	0,976	91,2 %	95,2 %	91 %
C) SVM	0,897	85,7 %	70 %	88,7 %
D) NEURAL NETWORK	0,987	93 %	97,8 %	92,9 %

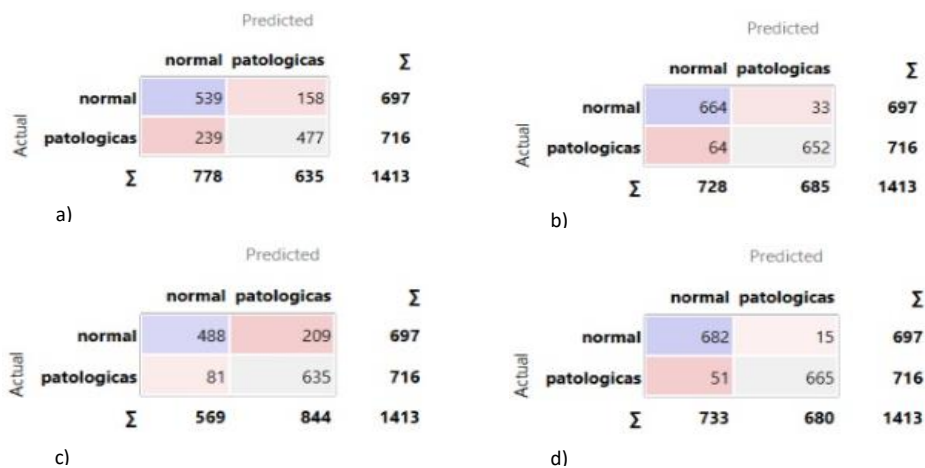


Figura 8 – Matrizes de confusão obtidas na análise PCA: a) kNN; b) Random Forest; c) SVM. D) Neural Network

Avaliando as matrizes de confusão obtidas dos classificadores no método PCA, verificamos que o Neural Network apresenta maior sensibilidade, sendo capaz de identificar verdadeiros positivos em 97,8% dos casos, e uma especificidade elevada, em que 93% das classificações são verdadeiros negativos, traduzindo-se numa precisão, de 92,9%, e numa maior AUC, de 0,987. Já o KNN é o pior classificador apresentando uma precisão mais baixa, de 69,2%, com sensibilidade de 77,3% e especificidade de 66,6%, e uma AUC de 0,781.

Tabela 1 - Resultados obtidos de AUC, Precisão, Sensibilidade e Especificidade em Rank

MODEL	AUC	PRECISÃO	SENSIBILIDADE	ESPECIFICIDADE
A) KNN	0,932	85,7 %	93,5 %	84,8 %
B) RANDOM FOREST	0,967	90 %	95,3 %	89,8 %
C) SVM	0,727	54,8 %	16,4 %	86,9 %
D) NEURAL NETWORK	0,972	89,5 %	98,7 %	88,7 %

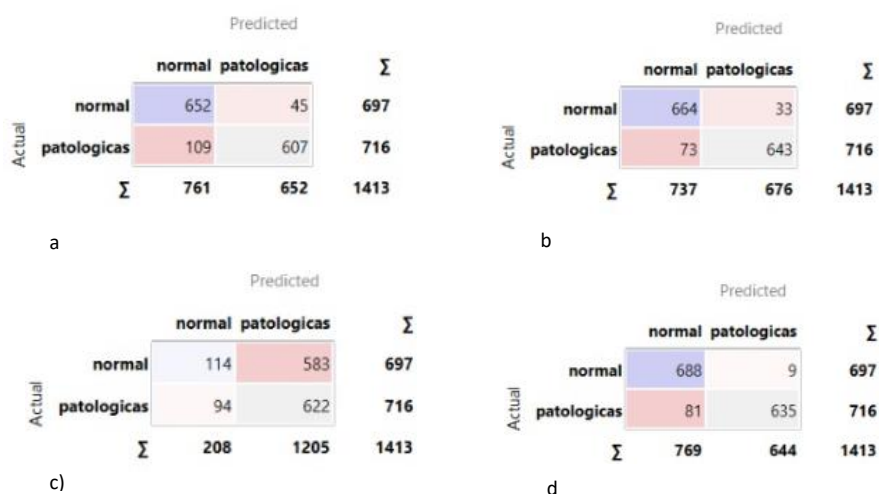


Figura 11 - Matrizes de confusão obtidas na análise Rank: a) kNN; b) Random Forest; c) SVM. d) Neural Network

Avaliando as matrizes de confusão obtidas dos classificadores no método Rank, verificamos que o Neural Network apresenta uma maior sensibilidade, sendo capaz de identificar verdadeiros positivos em 98,7% dos casos, e uma especificidade elevada, em que 88,7% das classificações são verdadeiros negativos, traduzindo-se numa precisão de 89,5%, com uma AUC de 0,972. Já o SVM é o pior classificador, apresentando uma precisão mais baixa, de 54,8%, com sensibilidade de 16,4% e especificidade de 86,9%, com uma AUC de 0,727.

Comparando os classificadores nos dois métodos, verificamos que o Neural Network tem melhor comportamento no método PCA. O classificador Random Forest tem valores de classificações semelhantes em ambos os métodos. O KNN apresenta melhor comportamento no método Rank, enquanto o SVM obtém melhores resultados no método PCA.

Verifica-se assim que com o método PCA, com redução para 5 componentes e 73% de variância explicada, o melhor classificador foi o *Neural Network*, com 0,987 de *Area Under the Curve* (AUC). No método Rank também o *Neural Network* foi o melhor classificador com 0,972 AUC, na qual se verificou uma sensibilidade de 97,8%, especificidade de 92,9% e precisão de 93%. Os resultados obtidos vão em linha com o observado na literatura, na qual a utilização de características *radiomics* para a classificação de radiografias de tórax tem mostrado resultados promissores.

Na ótica do técnico de radiologia é importante compreender que o potencial destas técnicas de processamento de exames radiológicos vão para além da radiografia do tórax, conceito enaltecido por Hardy M., (2020), referindo que estas técnicas de processamento já são uma realidade na prática clínica e que no futuro irão provocar alterações

disruptivas nos processos atualmente existentes. Como exemplo, verificamos aplicações a nível da RM cardíaca, onde Avendi *et al.* (2017) propuseram uma abordagem combinada de IA e um modelo deformável para a segmentação automatizada do ventrículo esquerdo, aplicações a nível da redução de dose em TC. Hwang *et al.* (2020) desenvolveram um método para a deteção e segmentação de nódulos pulmonares em exames de TC de baixa dose, enquanto Sharma *et al.* (2018) e Yao *et al.* (2019) discutiram este conceito a partir da radiografia do tórax. Por outro lado, Zhang *et al.* (2019) propuseram uma abordagem por redes neurais 3D para a classificação de imagens histopatológicas de cancro do pulmão, aliando o conceito de *radiomics* aos classificadores por redes neuronais como um conceito multidisciplinar e agregador de diferentes áreas médicas.

É importante referir que os métodos de IA são tão bons quanto a qualidade dos dados utilizados para treinar, ou como referenciado pelos cientistas de dados e engenheiros informáticos, o desempenho e validade dos métodos IA seguem o princípio “*garbage in*” e “*garbage out*”.²⁰ Assim é necessário reconhecer os desafios associados à implementação de estratégias de *radiomics* e IA na profissão de Técnicos de Radiologia, já que, em última instância será sobre o produto do nosso trabalho que estas metodologias se vão basear. É necessário abordar questões como a qualidade dos dados, a normalização, as preocupações com a privacidade e considerações éticas.

Tendo por base as preocupações referidas, Malamateniou *et al.*, 2021, apresentam algumas prioridades para a o desenvolvimento do técnico de radiologia nestas áreas, bem como integração orgânica da IA no ecossistema da Radiologia, nomeadamente:

1. Validação robusta das atuais ferramentas de IA em dados reais ainda não testados;
2. Estudos de investigação interdisciplinares mais prospetivos no domínio da Radiologia;
3. Regulamentação mais abrangente e amplamente aceite nos produtos IA para salvaguardar a sua utilização segura e eficaz;
4. Maior participação dos utilizadores dos serviços, incluindo os profissionais de saúde, doentes e respetivos prestadores de cuidados, na conceção e aplicação de ferramentas IA;
5. Esforços concertados com a indústria para encontrar soluções de IA explicáveis, a fim de permitir uma melhor compreensão e adoção para profissionais de saúde, bem como os utilizadores dos serviços;
6. Definição da responsabilização e de direito médico no caso de resultados erróneos resultantes da utilização de software e hardware baseado em IA;
7. Maior transparência nos processos de decisão para aumentar a confiança e ganhar aceitabilidade dos doentes;
8. Definição de competências profissionais claras e estratégias de desenvolvimento da extensão de funções para os profissionais de saúde, incluindo os técnicos de radiologia, num futuro em que a IA será fundamental.

As soluções de IA que proporcionam uma automatização semelhante da tomada de decisões para as tarefas radiográficas não são diferentes dos dispositivos radiológicos, na medida em que requerem provas clínicas e a aceitação e supervisão dos técnicos de radiologia antes de poderem ser amplamente implementadas. No entanto, um fator diferenciador fundamental é que os novos sistemas de IA têm o potencial de automatizar uma gama mais vasta de tarefas cognitivas de nível superior e, por conseguinte, pode argumentar-se que deve ser exigida uma maior diligência e provas antes da sua adoção.²¹

O estudo piloto que levou ao desenvolvimento deste artigo demonstra que é possível, através de processos educativos adequados (tanto a nível de pós-graduação, mestrado ou doutoramento), que os profissionais de saúde possam desenvolver, e mais importante, compreender, todo o processo metodológico que leva às soluções de IA que utilizamos já hoje nos serviços de Radiologia. É importante, conforme referido por Hardy, M., (2020), compreender que os técnicos de radiologia terão um papel cada vez mais importante na realização de auditoria e revisões regulares dos resultados e decisões de um sistema de avaliação de radiografias por IA. É provável a necessidade de um sistema “*peer-review*” ou de verificação pós-decisão dos sistemas automáticos baseados em IA, com o objetivo de estabelecer a sensibilidade, especificidade e exatidão do sistema, sendo são os técnicos de radiologia que devem aproveitar esta oportunidade para se apropriarem do processo e, talvez no futuro, estabelecerem sistemas de revisão e auditorias de IA na análise de exames radiológicos.

Conclusão

No estudo piloto realizado neste artigo, propusemos o desenvolvimento de uma metodologia de extração de características técnicas de *radiomics* e de classificação por métodos IA com base nos dados da ChestX-ray14, onde foram selecionadas 831 imagens patológicas e 813 imagens sem evidências de patologia. Dos diferentes métodos utilizados, verificou-se que o classificador Neural Network através do método PCA, apresentou uma precisão de 93%, com uma sensibilidade de 97,8% e uma especificidade de 92,9%.

Com base nestes resultados e na discussão associada ao papel do Técnico de Radiologia propomos como estudos futuros a aplicações destas metodologias na avaliação da qualidade técnica das radiografias presentes nas bases-de-dados públicas, mas também procurando desenvolver estratégias de integração destas tecnologias no dia-a-dia dos serviços de radiologia.

É essencial que os técnicos de radiologia possuam conhecimentos avançados sobre o funcionamento e a funcionalidade dos sistemas de IA, a fim de liderar e aproveitar as oportunidades proporcionadas por essa tecnologia. A necessidade de programas de formação e desenvolvimento profissional contínuo para apoiar a evolução dos técnicos de radiologia e sua adaptação aos novos serviços baseados em IA é de extrema importância e não pode ser negligenciada. A capacidade de aproveitar todo o potencial da IA na prática radiológica requer uma força de trabalho qualificada e capacitada.

Referências

1. Willemink, M. J., Persson, M., Pourmorteza, A., Pelc, N. J., & Fleischmann, D. Automatic exposure control in CT: The effect of patient size, anatomical region, and clinical task on tube current modulation. *Medical Physics*. 2020. 47(2), 643-658. doi:<https://doi.org/10.1007/s00330-014-3309-4>
2. van Timmeren, J., Cester, D., Tanadini-Lang, S. et al. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights Imaging* 11, 91 (2020). <https://doi.org/10.1186/s13244-020-00887-2>
3. Sanders VL, Flanagan J. Radiology Physician Extenders: A Literature Review of the History and Current Roles of Physician Extenders in Medical Imaging. *J Allied Health*. 2015 Winter;44(4):219-24. PMID: 26661701.
4. Miele, Vittorio, Chiara Andreoli, and Roberto Grassi. "The management of emergency radiology: key facts." *European journal of radiology* 59.3 (2006): 311-314. <https://doi.org/10.1016/j.ejrad.2006.04.020>
5. Kvak D, Chromcová A, Biroš M, Hrubý R, Kvaková K, Pajdaković M, Ovesná P. Chest X-ray Abnormality Detection by Using Artificial Intelligence: A Single-Site Retrospective Study of Deep Learning Model Performance. *BioMedInformatics*. 2023; 3(1):82-101. <https://doi.org/10.3390/biomedinformatics3010006>
6. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762. doi: <https://doi.org/10.1038/nrclinonc.2017.141>
7. van Leeuwen KG, de Rooij M, Schalekamp S, van Ginneken B, Rutten MJCM. How does artificial intelligence in radiology improve efficiency and health outcomes?. *Pediatr Radiol*. 2022;52(11):2087-2093. doi: <https://doi.org/10.1007/s00247-021-05114-8>
8. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*. 2017;318(22):2199-2210. doi: [10.1001/jama.2017.14585](https://doi.org/10.1001/jama.2017.14585)
9. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. doi:<https://doi.org/10.48550/arXiv.1702.05747>
10. Dyer T, Dillard L, Harrison M, et al. Diagnosis of normal chest radiographs using an autonomous deep-learning algorithm. *Clin Radiol*. 2021;76(6):473.e9-473.e15. doi: <https://doi.org/10.1016/j.crad.2021.01.015>
11. Chen B, Zhang R, Gan Y, Yang L, Li W. Development and clinical application of radiomics in lung cancer. *Radiat Oncol*. 2017;12(1):154. Published 2017 Sep 15. doi: <https://doi.org/10.1186/s13014-017-0885-x>
12. Homepage/Website: Kaggle.com [consultado 2022, 8 Nov]. Disponível em: <https://paperswithcode.com/dataset/chestx-ray14>
13. Homepage/Website: Radiopaedia.org [consultado em 2023, 29 Jan]. Radswiki T, Niknejad M, Suyama D, et al. Cardiomegaly. Reference article, Disponível em: <https://doi.org/10.53347/rID-12661>

14. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. 2016. *ArXiv:1612.07003*. <http://dx.doi.org/10.1148/radiol.2020191145>
15. Novikov AA, Lenis D, Major D. Fully Convolutional Architectures for Multi-Class Segmentation in Chest Radiographs. *ArXiv*.2018;1–13.<https://doi.org/10.48550/arXiv.1701.08816>
16. Prevedello LM, Halabi SS, Shih G, Wu CC, Kohli MD, Chokshi FH, et al. Challenges Related to Artificial Intelligence Research in Medical Imaging and the Importance of Image Analysis Competitions. *Radiol Artif Intell*. 2019;1(1):e180031. Doi: <https://doi.org/10.1148/ryai.2019180031>
17. Mladenic, D., Brank, J., Grobelnik, M., & Milic-Frayling, N. Feature selection using linear classifier weights. *SIGIR '04*.2004. <https://doi.org/10.1145/1008992.1009034>
18. Müller, A., & Guido, S. Introduction to Machine Learning with Python: A Guide for Data Scientists. *O'Reilly Media*.2016. ISBN 9352134575, 9789352134571, <http://cds.cern.ch/record/2229831>
19. Quintas J, Análise através da curva ROC: que ferramentas utilizar?. Dissertação de Mestrado em Bioinformática. Repositorium.sdum.uminho, 2020
20. Malamateniou, C., Knapp, K. M., Pergola, M., Hardy, M., & Hardy, M. A. Artificial intelligence in radiography: Where are we now and what does the future hold? *Radiography*. 2021. 27, S58–S62. <https://doi.org/10.1016/j.radi.2021.07.015>
21. Hardy, M., & Harvey, H. Artificial intelligence in diagnostic imaging: impact on the radiography profession. *British Journal of Radiology*. 2020. 93(1108), 20190840. <https://doi.org/10.1259/bjr.20190840>
22. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 2017. (pp. 2097-2106). <https://doi.org/10.48550/arXiv.1705.02315>
23. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Langlotz, C. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint arXiv*. 2017.1711.05225. <https://doi.org/10.48550/arXiv.1711.05225>
24. Yao, L., Zhang, C., Zhang, H., Chen, X., Zhou, H., & Zhang, D. A deep learning system for COVID-19 diagnosis from chest CT using weak label. *Medical Image Analysis*.2020. 67, 101851. <https://doi.org/10.1101/2020.03.12.20027185>
25. Avendi, M. R., Kheradvar, A., & Jafarkhani, H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Medical Image Analysis*, 2017. 30, 108-119. <https://doi.org/10.1016/j.media.2016.01.005>
26. Hwang, E. J., Park, S., Jin, K. N., Kim, J. I., Choi, S. H., & Kim, S. T. Deep learning-based fully automated detection and segmentation of pulmonary nodules on ultralow-dose CT: Retrospective cohort study. *Medical Imaging*.2020. 7(2), e15991. <https://doi.org/10.1016/j.compmedimag.2019.02.003>
27. Anis, Shazia, et al. "An overview of deep learning approaches in chest radiograph." *IEEE Access* 8 (2020): 182347-182354. <https://doi.org/10.1155/2021/5528144>
28. Zhang, Q., Wang, J., Wu, F., & Tang, L. Deep learning-based classification of lung cancer histopathology images using 3D convolutional neural networks. *Computers in Biology and Medicine*. 2019. 109, 157-166. <https://doi.org/10.1038%2Fs41598-021-84630-x>
29. Malamateniou, C., Knapp, K. M., Pergola, M., Hardy, M., & Hardy, M. A. Artificial intelligence in radiography: Where are we now and what does the future hold? *Radiography*. 2021. 27, S58–S62. <https://doi.org/10.1016/j.radi.2021.07.015>
30. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology*. 2018;286(3):800-809. doi:10.1148/radiol.2017171920
31. Lee, Seung Hak et al. "Radiomics in Breast Imaging from Techniques to Clinical Applications: A Review." *Korean journal of radiology* vol. 21,7 (2020): 779-792. doi: <https://doi.org/10.3348%2Fkjr.2019.0855>

Recebido /Received: 12/06/2023

Aceite / Accept: 10/07/2023