

Detecção de Doenças em Imagens de Raios-X da Coluna Lombo-Sacra com Convnets

Pablo Vieira^{1,2}, Luis Vogado^{1,2}, Lucas Lopes^{1,2}, Ricardo Lira^{1,2},
Pedro Santos Neto^{1,2}, Deborah Magalhães¹, Romuere Silva¹

¹Universidade Federal do Piauí – (UFPI) – Teresina – PI – Brazil
Programa de Pós-graduação em Ciência da Computação – Associação UFMA/UFPI

²Maida.health
Departamento de Pesquisa e Desenvolvimento – (P&D)

{pablovieira, ricardoalr, pasn, debora.vm, romuere}@ufpi.edu.br

{pablo.vieria, luis.vogado, lucas.lopes, pedro}@maida.health

Abstract. *Over the years, CAD systems to aid diagnosis have become more important. One exam that enables several CAD applications is the Lumbo-Sacral exam. This exam provides detailed radiographs of the spine, specifically of the lumbar, sacral, and coccygeal regions, allowing the detection of diseases such as arthrosis, scoliosis, spondylarthrosis, lordosis, osteophytes, disc space reduction, among others. In this context, we developed a CAD system methodology based on Deep-learning to detect anomalies in Lumbosacra. To this end, we used a dataset containing 16.024 exams, which is more heterogeneous than the state-of-the-art. In addition, we developed an ensemble in the classification using frontal and lateral images of the same exam, which allows classifying a greater number of pathologies, making the process more accurate, thus reducing false positives.*

Resumo. *Ao longo dos anos, o uso de sistemas CAD no auxílio a diagnósticos vem se tornando mais importante. Um exame que possibilita várias aplicações CAD é o de Lombo-Sacra. Esse exame fornece radiografias detalhadas da coluna vertebral, especificamente das regiões lombar, sacral e cocígea permitindo detectar doenças como artrose, escoliose, espondilartrose, lordose, osteófitos, redução do espaço discal, dentre outras. Nesse contexto, desenvolvemos uma metodologia de sistema CAD baseado em Deep-learning para atuar em Lombo-sacra. Para tal, utilizamos um conjunto de dados contendo 16,024 exames, mais heterogêneo que os do estado da arte. Além disso, desenvolvemos um ensemble na classificação utilizando imagens frontais e laterais do mesmo exame, o que permite classificar um número maior de patologias, o que torna o processo ainda mais preciso diminuindo assim os falsos positivos.*

1. Introdução

A Coluna Vertebral (CV) humana possui 33 vértebras intercaladas por 24 discos intervertebrais que, além da função estrutural e de movimentação, também desempenha o papel de canal para a medula espinhal. Esta estrutura é interconectada por ligamentos e

articulações, o que possibilitam sua flexibilidade. A coluna é dividida em quatro regiões distintas: região cervical, torácica, lombar, sacral e coccígea [Drake et al. 2019].

Alguns problemas de saúde como: artrose; escoliose; espinha bífida; espondilartrose; lordose; osteófitos; redução do espaço discal; dentre outros, estão associados a CV. Estes problemas podem causar dores e até mesmo a imobilização do paciente [Anouar 2019, O.A. 2019]. Exames de imagens médicas podem ser empregados na análise dos problemas na CV, sendo a radiografia o exame mais comum. Ela permite visualizar doenças degenerativas, fraturas, corpos estranhos dentre outras anormalidades [Food and Administration 2019].

A detecção de doenças em radiografias exige mão de obra qualificada, e por ser um trabalho repetitivo e meticuloso, pode levar à fadiga e consequentemente ocasionar erros (falsos positivos). Recentes sistemas *Computer-Aided Diagnosis* (CAD) vem utilizando técnicas de *Deep Learning* na detecção de doenças em radiografias. A adoção dos sistemas CAD visam auxiliar o trabalho dos especialistas e fornecer uma segunda opinião, aumentando a produtividade e diminuindo os falsos positivos, o que torna o processo mais eficiente [Esmail et al. 2020, Li et al. 2020, Zeng et al. 2020].

Neste trabalho, propomos uma metodologia de sistema CAD baseado em Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs) para classificar exames de lombo-sacra em normal e anormal. Para esse objetivo, apresentamos quatro contribuições: 1) utilizar uma nova base de dados, sendo esta maior e mais heterogêneo que as demais encontrados no estado da arte; 2) um pré-processamento para melhorar a qualidade das imagens de lombo-sacra; 3) um aumento de dados que leva em consideração as características do exame; e, 4) uma classificação de patologias em *ensemble* utilizando as imagens frontal e lateral de um mesmo exame, tornando o resultado mais confiável já que leva em consideração informações distintas contidas nas duas imagens.

2. Trabalhos Relacionados

Com a possibilidade de desenvolver um sistema CAD para a detecção de problemas na CV de pacientes, especificamente nas regiões lombo-sacra, apresentamos a Tabela 1 onde expomos as principais características de nosso levantamento sobre trabalhos relacionados. A seleção dos trabalhos foi realizada em três bases de dados distintas: Scopus, Web of Science e IEEE Xplore. As strings de busca utilizadas foram: "lumbo-sacral image classification", "x-ray image classification" e "lumbar spine with machine learning".

Ao analisar a Tabela 1, percebemos que o desenvolvimento de sistemas CAD para a CV possui dificuldades e lacunas que buscamos explorar neste trabalho como: a quantidade reduzida de exames; emprego somente de uma das imagens disponíveis no exame (frontal ou lateral); a não utilização de aumento de dados; a resolução de problemas específicos como, por exemplo, a escoliose, que diminuem o escopo de atuação; e, por último, a ausência da aplicação da técnica de *fine-tuning*.

Explorando os trabalhos apresentados na Tabela 1, nós desenvolvemos uma metodologia de sistemas CAD que visa classificar os exames de radiografias de lombo-sacra como normal e anormal, ou seja, todas as patologias presentes no conjunto de dados para emissão de alertas. Essa metodologia é possível pelo fato de utilizarmos um conjunto de dados bastante heterogêneo de radiografias de lombo-sacra com 16,024 exames. Para

Tabela 1. Resumo dos trabalhos levantados de nosso levantamento do estado da arte.

Trabalho	Exames	Tipo	Problema	DL	PP	AD	SEG	FT
[Jaremkó Jacob 2001]	40	frontal	escoliose	sim	sim	não	não	não
[Stanley Ronald 2001]	118	frontal	desvio	sim	sim	não	não	não
[Cherukuri et al. 2004]	714	lateral	osteoartrite, osteoporose e osteófitos	sim	sim	não	sim	não
[Veronezi et al. 2011]	466	lateral	osteoartrite primária	sim	não	não	não	não
[Tsantili-Kakoulidou Anna 2019]	595	frontal	escoliose	sim	não	não	sim	não
[Pan Yaling 2019]	248	frontal	escoliose	sim	sim	não	sim	não
[Weng et al. 2019]	1089	lateral	desvio sagital	sim	não	não	sim	não
[Lee et al. 2019]	334	lateral	densidade óssea	sim	não	não	não	não

DL: Deep-Learning; PP: Pré-processamento; AD: Aumento de Dados; SEG: Segmentação; FT: Fine-Tuning.

maximizar a quantidade de informações extraídas utilizamos tanto as imagens frontais quanto laterais em um *ensemble*, do mesmo exame já que cada uma permite a análise de uma perspectiva distinta da CV. Assim, fizemos também a transferência de aprendizado para empregarmos o *fine-tuning* dos modelos para cada contexto (frontal e lateral).

3. Metodologia Proposta

Nossa metodologia é exposta nas etapas: aquisição e pré-processamento das imagens de nosso conjunto de dados; classificação das imagens quanto ao tipo incidência (frontal, lateral e outros); segmentação do *token*; aumento de dados; treinamento com *fine-tuning*; e classificação em anormal e normal. A Figura 1 apresenta o fluxograma da metodologia.

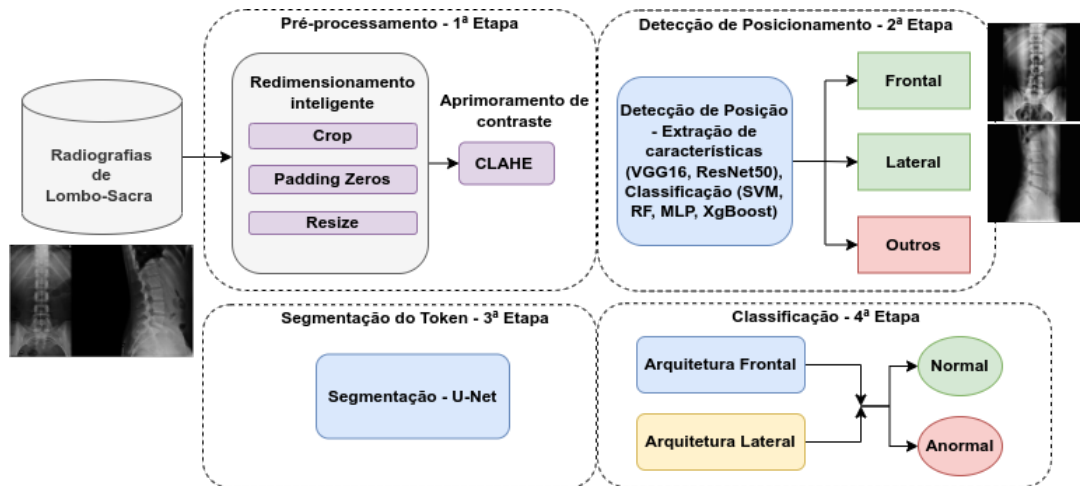


Figura 1. Etapas da metodologia. 1ª Aquisição dos exames de lombo-sacra; 2ª pré-processamento das imagens; 3ª detecção de posicionamento; 4ª segmentação de *tokens*; e 5ª classificação das imagens frontais e laterais em um *ensemble*.

3.1. Aquisição de Imagens

Um dos maiores desafios no desenvolvimento de sistemas CAD é a obtenção de conjuntos de dados robustos e heterogêneos, geralmente bases públicas podem ser limitadas nesses

aspectos. Sendo assim, nosso conjunto de dados possui 16,024 exames, sendo 8,265 classificados como normais e 7,759 anormais. Além disso, apresenta grande heterogeneidade por ser proveniente de diferentes hospitais, com características distintas na resolução das imagens, tipos de enquadramento, amostras de sexos e idades. A Figura 2 ilustra exemplos de imagens da incidência frontal e lateral pertencentes a um mesmo exame.

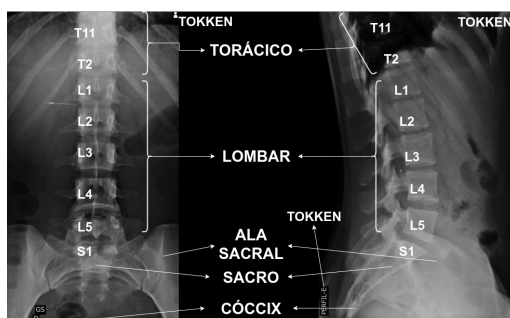


Figura 2. Exemplo de exame lombo-sacra com as imagens frontal e lateral do mesmo paciente. As legendas foram adicionadas para auxiliar a identificar características importantes.

Os exames desse conjunto foram obtidos em hospitais de todas regiões do Brasil, sendo classificados por especialistas em radiografias. A classe anormal contém problemas que a caracterizam como: desvio ou escoliose da coluna, artrose, espinha bífida, espondilartrose, lordose, osteófitos, redução do espaço discal, dentre outros. Ou seja, para a classe anormal utilizamos qualquer alteração ou patologia apontada pelo especialista.

3.2. Pré-processamento

Para tornar o processo de classificação mais eficaz realizamos uma etapa de pré-processamento que busca melhorar as características das imagens. A Figura 3 ilustra os efeitos do pré-processamento nas imagens originais.

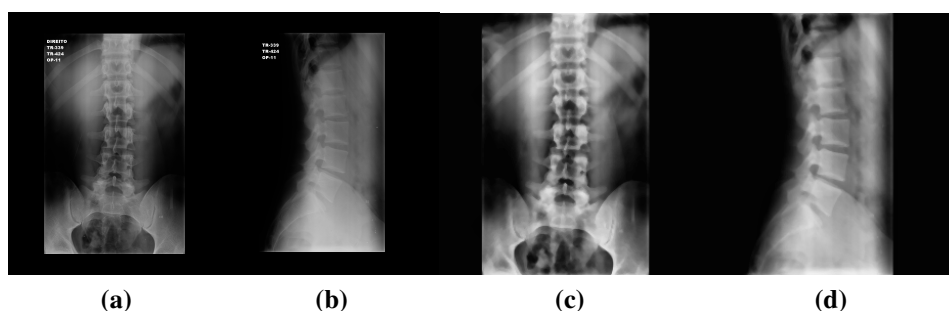


Figura 3. Exemplos do pré-processamento em imagens do nosso conjunto de dados: (a) Frontal original, (b) Lateral original, (c) Frontal pré-processada, (d) Lateral pré-processada.

Nosso pré-processamento consiste em: aplicação do limiar de Otsu [Otsu 1979], onde removemos a região de fundo da imagem, restando somente a região de interesse; transformação da imagem em um quadrado adicionando bordas de pixels zero, pois esse é o formato padrão adotado por CNNs. Finalmente, a imagem quadrada é redimensionada para as dimensões 256×256 , sem que ocorram distorções.

Nas radiografias, os ossos e tecidos menos densos possuem fronteiras lisas, o que pode dificultar a detecção de anomalias [Vieira et al. 2021a]. Para superar essa limitação,

melhoramos o contraste das imagens aplicando o *Limited Contrast Adaptive Histogram Equalization* (CLAHE) que divide as imagens em blocos, que são equalizados pelo histograma dessas regiões [Pizer et al. 1990]. Nas Figuras 3c e 3d observamos que o contraste aperfeiçoado pelo CLAHE evidencia as vértebras nas radiografias.

De acordo com [Zech et al. 2018] e [Geirhos et al. 2020], a presença de *tokens* metálicos de hospitais e meta dados, destacados na Figura 2, em radiografias pode enviesar o aprendizado das CNNs. Para evitar que o aprendizado seja comprometido por esses fatores, propomos segmentar essas regiões com a CNN U-Net. O emprego de uma CNN nesse processo se deu por sua capacidade de generalização, visto que nosso conjunto de dados contém padrões variados quanto ao tamanho, localização e disposição dos *tokens*.

Para a remoção dos *tokens*, especialistas marcaram 500 imagens frontais e 500 laterais onde usamos 80% e 20% para treinar e validar a segmentação com a U-Net. Após a remoção do *token* com a U-Net, usamos o *Fast Marching Method* (FMM) proposto em [Telea 2004] para corrigir a área segmentada. Esse método leva em consideração informações dos pixels vizinhos, partindo sempre das bordas da região de interesse. Os pixels da região são substituídos pela soma ponderada e normalizada de todos os pixels da vizinhança. As Figuras 3a e 3c; 3b e 3d mostram exemplos de imagens antes e após a eliminação dessas regiões.

3.3. Aumento de Dados

Apesar de nosso conjunto de dados possuir quantidade considerável de amostras, o emprego do Aumento de Dados (AD) torna ele ainda mais robusto diminuindo a possibilidade de *overfitting* [Shorten Connor 2019]. Durante o treinamento empregamos o AD em tempo de execução, onde as operações de aumento foram aplicadas de forma aleatória misturando-as entre si. Assim, a cada época, devido o AD são apresentadas ao modelo "novas" imagens. Neste trabalho, nós utilizamos operações matemáticas que preservam características importantes nas imagens, simulando alterações de posições e qualidade de forma randômica. As operações foram rotação variando de 0° a 8°; zoom variando de 0% a 3%; deslocamento da imagem no sentido da altura e largura no intervalo de 0,1; brilho variando de 0% a 20%; ruído Gaussiano variando de 0% a 8%.

3.4. Redes Neurais Convolucionais

No levantamento do estado da arte, na Tabela 2, observamos que as Redes Neurais Convolucionais são promissoras para soluções CAD voltadas para radiografias da coluna. Sendo assim optamos por utilizar neste trabalho as arquiteturas: VGG16 [Simonyan and Zisserman 2014], VGG19 [Simonyan and Zisserman 2015] e ResNet50 [He et al. 2016] nas etapas de extração de características e classificação. A escolha dessas arquiteturas se baseou em suas capacidades de convergência em variados tipos de imagens médicas [Vieira et al. 2021b].

3.5. Transferência de Aprendizado

Os autores de [Paras. 2017] observaram que a transferência de aprendizado usando redes pré-treinadas em imagens não médicas, como o ImageNet [Russakovsky et al. 2015], superou redes treinadas do zero em classificações de radiografias. Existem duas soluções baseadas na transferência de aprendizado, são elas: extração de características e *fine-tuning*. Para o desenvolvimento do sistema CAD, empregamos a extração de características na detecção de incidências e o *fine-tuning* na classificação binária de exames.

3.6. Extração de Características para Classificação

A extração de características utiliza CNNs pré treinadas na extração de vetores de características das imagens, que são utilizados no treino de classificadores como *Support Vector Machines* (SVM) [Cortes and Vapnik 1995], *Random Forest* (RF) [Breiman 2001], *Multilayer Perceptron* (MLP) [Popescu et al. 2009] e *Scalable Tree Boosting System* (Xg-Boost) [Chen and Guestrin 2016]. Essa solução foi empregada na detecção de incidências (frontal, lateral e outros). As CNNs empregadas nessa etapa foram a VGG16 e ResNet50.

3.7. Fine-Tuning

Em [Tajbakhsh et al. 2016] observaram que uma CNN pré-treinada com *fine-tuning* (FT) supera uma treinada do zero na classificação de radiografias. Para realizar FT, geralmente empregamos CNNs pré-treinadas no conjunto de dados ImageNet, onde continuamos seu treinamento no conjunto de dados que tentamos resolver. Para este estudo, optamos pelo FT Profundo, onde todas as camadas da CNN são liberadas para continuar o aprendizado no novo conjunto de dados [Vieira et al. 2021a], essa escolha se deu devido ao conjunto de dados possuir quantidade de dados suficiente para o aprendizado total da rede.

Para o FT removemos das CNNs de classificação (frontais e laterais), as camadas totalmente conectadas, em seguida, adicionamos 3 novas logo após as camadas convolucionais. Na camada de entrada utilizamos 1,024 neurônios, 128 na camada oculta e 2 na de saída, esses valores foram selecionados por experimentos definidos empiricamente. No treinamento, utilizamos 24 épocas com taxa de aprendizado de 0,0001. O algoritmo de otimização utilizado foi *A Method for Stochastic Optimization* (ADAM) [Kingma and Ba 2017] com decaimento de 0,0001. A função de custo utilizada foi a *categorical crossentropy* com função de ativação *Softmax* [Goodfellow et al. 2016].

3.8. Ensemble

Ao pegarmos modelos que convergiram para mínimos locais e aplicarmos a média, podemos construir uma estimativa melhorada através do *ensemble*. Uma forma de entender esse fato é considerar que, em geral, as redes que caíram em diferentes mínimos locais terão um desempenho ruim em diferentes regiões do espaço de características e, portanto, os fatores que geram o erro não serão fortemente correlacionados. É essa falta de correlação que impulsiona o método de média do *ensemble* [Perrone and Cooper 1993].

3.9. Métricas de Validação

Para avaliar a metodologia proposta, selecionamos métricas frequentemente utilizadas na literatura: acurácia (Acc), precisão (P), especificidade (E), sensibilidade (S), kappa (K), F1-Score (F1), área sobre curva ROC (AUC) [Goodfellow et al. 2016].

4. Experimentos

Nós projetamos experimentos a fim de validar a detecção de incidências e classificação de anormalidades, que são detalhados a seguir, e na Tabela 2.

Detecção de Incidências: Uma etapa importante de nossa metodologia é detectar imagens corretas, onde procuramos três tipos de imagens (frontal, lateral e outros), o tipo outros pode ser qualquer uma que não se enquadre nas duas primeiras classes.

Classificação de Anormalidades: Treinamos modelos distintos um *ensemble* com as imagens frontais e laterais, como ilustra a Figura 4 por um \oplus das arquiteturas especializadas. Isso potencializa nossa metodologia, já que algumas doenças são mais facilmente detectáveis em um desses tipos de imagens.

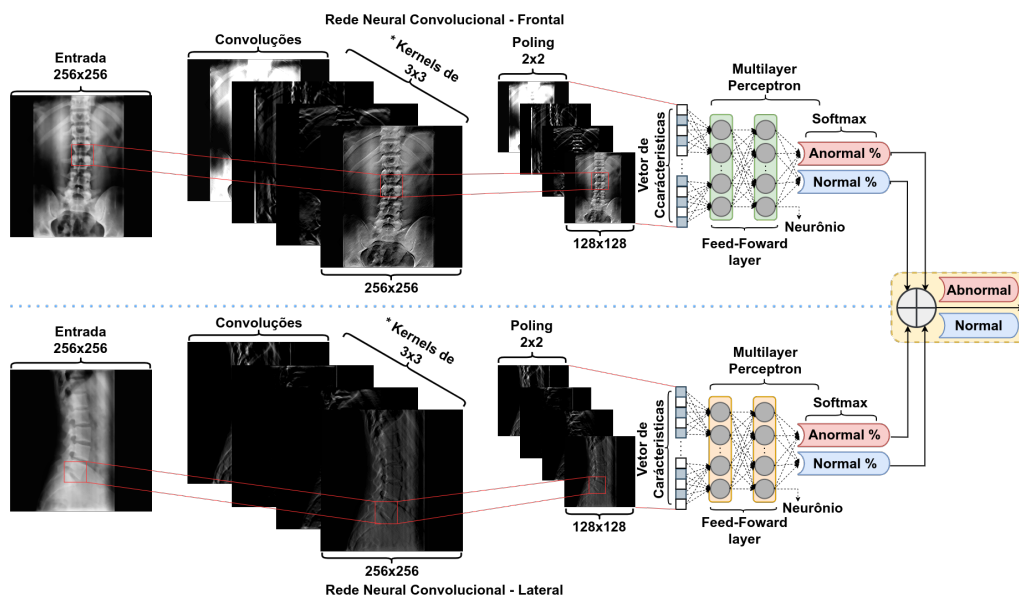


Figura 4. Ilustração de modelos de CNNs distintos para as incidências frontal e lateral.

Tabela 2. Tabela com o resumo dos experimentos realizados em nossa pesquisa.

Classes	Amostras	Método
	Detecção de Incidências.	
[frontal, lateral, cóccix, outros]	[1,278, 1,262, 1,220, 661]	[Seção 3.5]
	Classificações Frontais e Laterais.	
[normal, anormal]	[8,265, 7,759]	[Seção 3.7]
[normal, anormal]	[4,183, 4,454]	[Seção 3.7]

5. Resultados

A Tabela 3 mostra resultados dos experimentos de detecção de posicionamento, a metodologia empregada aqui obteve resultados promissores. A ResNet50 superou todos os resultados da VGG16, entendemos que isso se deu por sua maior profundidade que extraiu características mais profundas, e aliada aos *kernels* do SVM que conseguem lidar com dados de grandes dimensões conseguimos 100% de acerto. Também é perceptível visualmente que o problema não é tão complexo de resolução.

Na etapa de classificação de anormalidades, empregamos imagens frontais e laterais, para cada tipo de imagem treinamos modelos distintos das CNNs VGG16, VGG19 e ResNet50. Os resultados destes experimentos são apresentados na Tabela 4, e consideramos promissores já que esse problema nunca foi resolvido antes e a taxa de acc está acima dos 80% para quase todas arquiteturas, indicando que ocorreu boa generalização. Podemos observar que as arquiteturas mais rasas, VGG16 e VGG19 obtiveram os melhores resultados, o fato de possuírem menor complexidade que a ResNet50 diminuiu a

Tabela 3. Tabela de resultados obtidos na detecção de posicionamento.

CNN	Classifier	Acc	Kappa	Preci.	F1-Score	AUC
VGG16	RF	0,923	0,896	0,910	0,918	0,941
VGG16	SVM	0,973	0,963	0,970	0,972	0,980
VGG16	MLP	0,982	0,976	0,979	0,982	0,987
VGG16	XgBoost	0,966	0,954	0,960	0,964	0,974
ResNet50	RF	0,973	0,963	0,965	0,969	0,978
ResNet50	SVM	1,0	1,0	1,0	1,0	1,0
ResNet50	MLP	0,993	0,991	0,992	0,993	0,995
ResNet50	XgBoost	0,986	0,981	0,984	0,985	0,990

Em **negrito** os melhores resultados.

probabilidade do *overfitting*, o que melhorou sua generalização. Por fim, realizado o teste T de *Student* [Gibson et al. 2017] para comparar os resultados estatisticamente a um nível de significância de 5%. O teste evidenciou que há diferença entre os principais resultados apresentados.

Tabela 4. Tabela de resultados obtidos na classificação anormal e normal (frontal e lateral).

CNN	Acc	P	E	S	K	F1	AUC
Experimentos com Imagens Frontais.							
VGG16	0,825	0,841	0,842	0,808	0,650	0,824	0,825
VGG19	0,829	0,820	0,808	0,849	0,658	0,835	0,829
ResNet50	0,799	0,782	0,760	0,836	0,597	0,808	0,798
Experimentos com Imagens Laterais.							
VGG16	0,865	0,887	0,890	0,840	0,729	0,863	0,865
VGG19	0,851	0,861	0,860	0,843	0,703	0,852	0,852
ResNet50	0,850	0,844	0,835	0,865	0,700	0,854	0,850

Em **negrito** os melhores resultados.

A Tabela 5 apresenta os resultados da combinação entre as CNNs avaliadas na Tabela 4. Os resultados foram ordenados de acordo com a AUC e Acc obtidas por cada combinação de CNN frontal e lateral. O melhor resultado obtido foi com as CNNs VGG19 e VGG16 com AUC de 0.862 e Acc de 0.861. Apesar do melhor resultado, a arquitetura frontal da VGG19 quando combinada com as demais CNNs laterais apresenta os três melhores resultados das combinações apresentadas.

Tabela 5. Tabela de resultados obtidos na classificação como anormal e normal em *ensemble*.

CNN Frontal	CNN Lateral	Acc	P	E	S	K	F1	AUC
VGG19	VGG16	0,861	0,882	0,884	0,840	0,723	0,860	0,862
VGG19	VGG19	0,860	0,881	0,884	0,838	0,722	0,860	0,861
VGG19	ResNet50	0,856	0,874	0,875	0,838	0,713	0,860	0,857
VGG16	ResNet50	0,853	0,897	0,905	0,802	0,706	0,847	0,853
VGG16	VGG16	0,851	0,897	0,905	0,799	0,703	0,845	0,852
VGG16	VGG19	0,851	0,897	0,905	0,798	0,702	0,845	0,852
ResNet50	VGG16	0,845	0,863	0,865	0,826	0,691	0,844	0,846
ResNet50	VGG19	0,844	0,860	0,861	0,827	0,688	0,843	0,844
ResNet50	ResNet50	0,842	0,857	0,857	0,827	0,684	0,842	0,842

Em **negrito** os melhores resultados.

6. Discussões

Apesar de resultados robustos, as CNNs são conhecidas como caixas pretas já que é difícil explicar como os conseguem. No entanto, [Selvaraju et al. 2016] desenvolveram um método que auxilia na explicação, pois possibilita visualizar regiões importantes para a classificação das CNNs o *Gradient-weighted Class Activation Mapping* (Grad-CAM). A Figura 5 apresentam exemplos de regiões importantes para a classificação das CNNs.

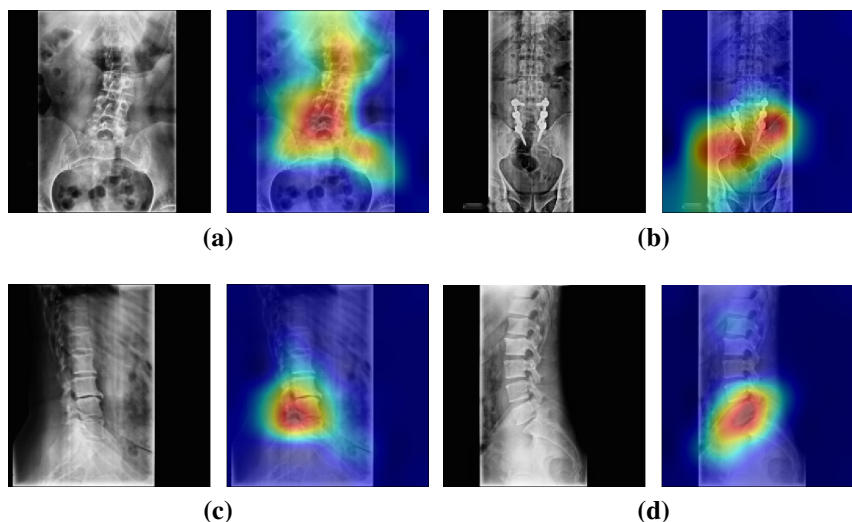


Figura 5. Exemplos de Grad-CAM das imagens frontais e laterais.

Na Figura 5a, observamos como a CNN interpretou uma imagem frontal, encontrando o que o médico laudou como: *contendo desvio lombar para a direita e contendo redução de espaço discal L5-S1*. Na Figura 5b observamos que a CNN mais uma vez encontrou o que foi laudo pelo médico: *O exame de controle pós-operatório de artrodese em coluna lombar com fixação com placa metálica e parafusos em L4 a S1; osteófitos anteriores incipientes; redução do espaço discal de L4 a S1; e esclerose de interapofisárias de L5-S1*. Na Figura 5c o Grad-CAM de uma imagem lateral aponta para o laudo médico: *escoliose lombar de convexidade a esquerda; osteófitos em vários corpos vertebrais; e várias reduções do espaço discais*. Na Figura 5d vemos a precisão da CNN em detectar o laudo apontado pelo médico: *osteófitos em alguns corpos vertebrais; algumas reduções do espaços discal*.

Nas classificações, especificamente frontal notamos a dificuldade das CNNs em conseguir bons resultados. Para entendermos o motivo, analisamos junto a especialistas em radiografias quais possíveis motivos. A análise mostrou que algumas doenças eram mais perceptíveis em imagens laterais como, por exemplo, as doenças de osteófitos, lordose e redução do disco nas vértebras L5-S1. Então, em conjunto com os especialistas, decidimos remover essas imagens de treino da classificação frontal. Assim, o classificador das imagens laterais ficaria responsável por denunciar esses problemas no *ensemble*.

A medicina não é uma ciência exata, alguns exames podem levar a conclusões diferentes por especialistas distintos. Percebemos isso na classificação de desvio na coluna. Nossos dados são de hospitais de todas regiões do Brasil, assim encontramos algumas diferenças. Algumas imagens que as CNNs classificavam como contendo desvio, eram

laudadas por médicos como normais (o contrário também ocorria). Isso se dava ao treinarmos a CNN com uma imagem que apresentava características de uma classe, no entanto, outra imagem com características similares pertencia a outra classe, o que dificultava o aprendizado da arquitetura. Para resolver este problema, refizemos os laudos destes exames com auxílio de especialistas em radiografias.

7. Conclusão

Neste trabalho, apresentamos uma metodologia CAD completa para auxiliar especialistas a laudar exames lombo-sacra em radiografias. Para tal, utilizamos um conjunto de dados robusto com 16,024 exames de variados gêneros e idades de pacientes e oriundos de diferentes hospitais e equipamentos, sendo maior que os demais encontrados em nosso levantamento na literatura. Implementamos etapas de pré-processamento para melhorar a qualidade das imagens. Segmentamos *tokens* que poderiam enviesar o aprendizado das CNNs. Desenvolvemos um aumento de dados, que leva em consideração as características do exame, para aprimorar nosso conjunto de dados. Ainda, utilizamos arquiteturas auxiliares para detecção de incidências, etapas de triagem importantes de uma solução CAD.

Em nossas classificações (anormal e normal) geramos duas CNNs distintas treinadas nas imagens frontais e laterais. Observamos resultados promissores, principalmente referente a arquitetura VGG16, em ambos os contextos. Percebemos que *ensemble* tornou nossas classificações ainda mais robustas, já que cada tipo de imagem pode fornecer informações diferentes do mesmo paciente. Com isso, nós acreditamos que podemos auxiliar especialistas na tomada de decisões em laudos de exames de lombo-sacra, podendo tornar o processo mais ágil, eficiente e consequentemente diminuindo os falsos positivos.

Para trabalhos futuros, nos pretendemos testar novas arquiteturas de CNNs. E, por fim, pretendemos desenvolver uma metodologia de busca por melhor conjunto de hiper-parâmetros de configuração das arquiteturas no ajuste-fino.

Referências

- Anouar, O. I. B. P. L. C. C. D. B. L. B. (2019). Classification of coronal imbalance in adult scoliosis and spine deformity: a treatment-oriented guideline. *European Spine Journal*.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Chen, T. and Guestrin, C. (2016). Xgboost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Cherukuri, M., Stanley, R., Long, R., Antani, S., and Thoma, G. (2004). Anterior osteophyte discrimination in lumbar vertebrae using size-invariant features. *Computerized Medical Imaging and Graphics*, 28(1):99 – 108.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.
- Drake, R., Vogl, W., and Mitchell, A. (2019). *Gray’s Anatomy for Students*. Gray’s Anatomy Series. Elsevier.
- Esmail, K. M., El-Din, H. E., and A., S. M. (2020). Cascaded deep learning classifiers for computer-aided diagnosis of covid-19 and pneumonia diseases in x-ray scans. *Complex Intelligent Systems*.

- Food, U. and Administration, D. (2019). Medical x-ray imaging. <https://www.fda.gov/radiation-emitting-products/medical-imaging/medical-x-ray-imaging>.
- Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., and Wichmann, F. A. (2020). Shortcut learning in deep neural networks.
- Gibson, E., Hu, Y., Huisman, H. J., and Barratt, D. C. (2017). Designing image segmentation studies: Statistical power, sample size and reference standard quality. *Medical Image Analysis*, 42:44–59.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.
- Jaremko Jacob, Poncet Philippe, R. J. H. J. D. J. L. H. Z. R. (2001). Estimation of spinal deformity in scoliosis from torso surface cross sections. *Spine*, 26:1583–1591.
- Kingma, D. P. and Ba, J. (2017). Adam: A method for stochastic optimization.
- Lee, S., Choe, E., Kang, H., Yoon, J., and Kim, H. (2019). The exploration of feature extraction and machine learning for predicting bone density from simple spine x-ray images in a korean population. *Skeletal Radiology*, 49.
- Li, X., Shen, L., Xie, X., Huang, S., Xie, Z., Hong, X., and Yu, J. (2020). Multi-resolution convolutional networks for chest x-ray radiograph based lung nodule detection. *Artificial Intelligence in Medicine*, 103:101744.
- O.A., Y. A. D. A. P. (2019). Spondyloarthrosis: pathogenesis, clinic, diagnosis and treatment (literature review and own experience). *Journal of Clinical Practice*, 10(4):61–73.
- Otsu, N. (1979). A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66.
- Pan Yaling, Chen Qiaoran, C. T. W. H. Z. X. F. Z. L. Y. (2019). Evaluation of a computer-aided method for measuring the cobb angle on chest x-rays. *European Spine Journal*, 28.
- Paras., L. (2017). Deep Convolutional Neural Networks for Endotracheal Tube Position and X-ray Image Classification: Challenges and Opportunities. *Journal of Digital Imaging*.
- Perrone, M. and Cooper, L. (1993). When networks disagree: Ensemble methods for hybrid neural networks. *Neural networks for speech and image processing*.
- Pizer, S. M., Johnston, R. E., Ericksen, J. P., Yankaskas, B. C., and Muller, K. E. (1990). Contrast-limited adaptive histogram equalization: Speed and effectiveness. *IEEE*.
- Popescu, M.-C., Balas, V. E., Perescu-Popescu, L., and Mastorakis, N. (2009). Multi-layer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7):579–588.

- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.
- Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., and Batra, D. (2016). Grad-cam: Why did you say that?
- Shorten Connor, K. T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*.
- Stanley Ronald, L. R. (2001). A radius of curvature-based approach to cervical spine vertebra image analysis. *Biomedical sciences instrumentation*, 37:385–90.
- Tajbakhsh, N., Shin, J., Gurudu, S., Hurst, R., Kendall, C., Gotway, M., and Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312.
- Telea, A. (2004). An image inpainting technique based on the fast marching method. *Journal of Graphics Tools*, 9(1):23–34.
- Tsantili-Kakoulidou Anna, Horng Ming-Huwi, K. C.-P. F. M.-J. L. C.-J. S. Y.-N. (2019). Cobb angle measurement of spine from x-ray images using convolutional neural network. *Computational and Mathematical Methods in Medicine*, pages 195 – 199.
- Veronezi, C. C. D., SimÃ, P. W. T. d. A., Santos, R. L. d., Rocha, E. L. d., MelÃ, S., Mattos, M. C. A. d., and Cechinel, C. (2011). Computational analysis based on artificial neural networks for aiding in diagnosing osteoarthritis of the lumbar spine. *Revista Brasileira de Ortopedia*, 46:195 – 199.
- Vieira, P., Sousa, O., Magalhães, D., Rabêlo, R., and Silva, R. (2021a). Detecting pulmonary diseases using deep features in x-ray images. *Pattern Recognition*, 119:108081.
- Vieira, P. A., Magalhães, D. M., Carvalho-Filho, A. O., Veras, R. M., Rabêlo, R. A., and Silva, R. R. (2021b). Classification of covid-19 in x-ray images with genetic fine-tuning. *Computers Electrical Engineering*, 96:107467.
- Weng, C.-H., Wang, C.-L., Huang, Y.-J., Yeh, Y.-C., Fu, C.-J., Yeh, C.-Y., and Tsai, T.-T. (2019). Artificial intelligence for automatic measurement of sagittal vertical axis using resunet framework. *Journal of Clinical Medicine*, 8:1826.
- Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., and Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine*, 15(11):1–17.
- Zeng, Y., Liu, X., Xiao, N., Li, Y., Jiang, Y., Feng, J., and Guo, S. (2020). Automatic diagnosis based on spatial information fusion feature for intracranial aneurysm. *IEEE Transactions on Medical Imaging*, 39(5):1448–1458.