

# Análise da influência do tamanho da região de interesse no diagnóstico auxiliado por computador de fraturas por fragilidade óssea

Thiago F. C. da Fonseca\*, Carolina Y. V. Watanabe\*, Jamilly G. Maciel<sup>‡</sup>, Marcello H. Nogueira-Barbosa<sup>‡</sup>, Caetano Jr. Traina<sup>†</sup>, Agma J. M. Traina<sup>†</sup> e Jonathan S. Ramos\*

\* Departamento Acadêmico de Ciência da Computação, Universidade Federal de Rondônia, Porto Velho – RO (Brasil),  
Email: thiago.coucello@gmail.com, {carolina,jonathan}@unir.br

<sup>†</sup> Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, – SP (Brasil)

<sup>‡</sup> Hospital das Clínicas da Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto – SP (Brasil)

**Abstract**—O diagnóstico de fraturas lombares devido a osteoporose e osteopenia é uma tarefa desafiadora devido a semelhança entre os dois casos. A princípio, ambas são assintomáticas, podendo posteriormente levar a fratura por fragilidade óssea. Na literatura há estudos promissores na diferenciação entre corpos vertebrais com e sem fraturas. Nestes estudos são usados a abordagem radiômica e redes neurais com transferência de aprendizado em corpos vertebrais segmentados. Contudo, a segmentação manual ou semi-automática ainda é um processo trabalhoso. Por este motivo, neste artigo é realizada a comparação da influência do tamanho da região de interesse sobre a classificação final. Assim, nosso principal objetivo consiste o quanto de detalhe é necessário na segmentação de forma a se obter resultados aceitáveis. Desta forma, é possível que o processo de segmentação seja facilitado ou automatizado em trabalhos futuros. Os resultados experimentais e a validação estatística obtidos com uma base de imagens representativa contendo 2400 casos, divididos meio a meio entre classes com e sem fraturas, demonstram que é possível reduzir o tamanho das regiões de interesse em até 45% sem diferença significativa na acurácia. Com o modelo Xception apresentando resultados mais próximos quando comparado com as imagens originais.

**Abstract**—The diagnosis of lumbar fractures due to osteoporosis and osteopenia is a challenging task due to the similarity between the two cases. At first, both are asymptomatic, and may subsequently lead to fracture due to bone fragility. In the literature, there are promising studies on the differentiation between vertebral bodies with and without fractures. In these studies, the radiomics approach and neural networks with transfer learning in segmented vertebral bodies are used. However, manual or semi-automatic segmentation is still a laborious process. For this reason, this article compares the influence of the size of the region of interest on the final classification. Thus, our main objective is how much detail is needed in the segmentation in order to obtain acceptable results. In this way, it is possible that the segmentation process will be facilitated or automated in future works. The experimental results and the statistical validation obtained with a representative image database containing 2400 cases, divided half and half between classes with and without fractures, demonstrate that it is possible to reduce the size of the regions of interest by up to 45% without significant difference in accuracy. With the Xception model showing closer results when compared to the original images.

## I. INTRODUÇÃO

A osteoporose e osteopenia são doenças que ocorrem de forma gradativa nos organismos e reduzem a Densidade Mineral Óssea (do inglês, *Bone Mineral Density* - BMD). Essa redução pode ocasionar fraturas por fragilidade, geralmente na coluna lombar (do inglês, *Vertebral Fragility Fractures* - VFF) [1], [2]. Em 2022, as pesquisas da Fundação Internacional de Osteoporose<sup>1</sup> constataram que, globalmente, 20% dos homens e 60% das mulheres acima dos 50 anos de idade sofrerão algum tipo de VFF [2]. As VFFs podem causar a redução da função física, imobilidade, isolamento social e uma possível depressão, reduzindo a qualidade de vida [2], [3], tornando-a uma questão de saúde pública.

Atualmente o padrão para diagnóstico e prevenção de VFFs resume-se na análise da BMD a partir de exames de DEXA (do inglês, *Dual X-Ray Absorptiometry*) [4]. No entanto, esse método apresenta uma sensibilidade de apenas 50% [5]. Além de ser questionável já que as VFFs podem ocorrer em pacientes com uma BMD normal.

Estudos da literatura apresentam resultados promissores quanto a aplicação de Inteligência Artificial, tanto com modelos de Aprendizado de Máquina (AM) tradicionais quanto em redes neurais profundas (do inglês, *Deep Neural Network* - DNN). Nestes estudos foram utilizadas imagens de ressonância magnética (do inglês, *Magnetic Resonance Imaging* - MRI) das quais foram extraídas características radiômicas de textura e demonstrada a forte correlação com a BMD, que foi confirmada em outros trabalhos [2], [4], [6]–[11]. Os autores reportaram 89.8% de AUC usando transferência de aprendizado com a rede ResNet101V2 e segmentação semi-automática [2]. Contudo, em sistemas de diagnóstico auxiliado por computador (CAD), a segmentação ainda é um passo laborioso, mesmo com a existência de métodos semi-automáticos [11]–[15].

Neste trabalho, foi analisado a influência do tamanho das regiões de interesse (do inglês, *Region Of Interest* – ROI) no desempenho das DNNs. Os resultados obtidos demonstram

<sup>1</sup><https://www.osteoporosis.foundation/>

que é possível reduzi-las em até 45% sem diferenças significativas em relação ao uso da imagem original (Tabela I), mostrando que é possível reduzir a precisão nos processos de segmentações manuais ou semi-automáticas sem perdas consideráveis em desempenho, reduzindo também a carga de trabalho e o tempo consumido nessas etapas. O modelo Xception obteve resultados que mais se aproximaram do equivalente ao uso das imagens originais. Este trabalho está organizado da seguinte forma: a Seção II apresenta os materiais e métodos. A Seção III reporta os resultados, experimentos e discussão. Finalmente, a Seção IV conclui nosso trabalho.

## II. MATERIAIS E MÉTODOS

O trabalho foi dividido em duas etapas, primeiramente a aquisição e pré-processamento de uma base de imagens que apresentasse variações no tamanho das ROIs, posteriormente o treinamento de modelos de DNN para análise do impacto da nova base de imagens em seus desempenhos.

### A. Base de imagens

A base de imagens foi disponibilizada por [2] e é composta de 94 sequências de MRI com ponderações T1 e T2 (T1w e T2w) obtidas a partir de 47 mulheres voluntárias na pós-menopausa. Os exames foram adquiridos por um sistema Tesla 3.0 (Philips Achieva, Philips Medical Systems, Best, the Netherlands).

Para a variação do tamanho das ROIs a área efetiva dessas regiões foi reduzida. Cada imagem teve sua área reduzida às porcentagens no intervalo de 5% a 100% da área original com incrementos de 5%, totalizando em 20 variações.

O processo de redução da área foi realizado utilizando morfologia matemática, onde as bordas da ROI original eram detectadas e retiradas até alcançar a porcentagem alvo. Caso a remoção ultrapasse o alvo, apenas a quantidade de pixels necessária é removida. Para melhor ilustrar, a Figura 1 apresenta, respectivamente, as ROIs geradas com 90%, 60%, 30% e 10% da área (em ciano) em relação ao padrão-ouro (em vermelho). Porém, na prática, as imagens alimentadas aos modelos contém apenas a área em ciano sendo o restante substituído por um fundo preto.

A base de imagens pós-processamento conta com 2.400 imagens por porcentagem calculada divididas igualmente entre imagens contendo e não contendo VFFs.

### B. Treinamento e análise dos resultados

Para a etapa de treinamento foram utilizados os seguintes modelos pré-treinados no dataset da ImageNet<sup>2</sup> Xception, ResNet101V2, DenseNet201, e MobileNet, os quais apresentaram os melhores resultados em [2]. Também foram utilizadas as técnicas de congelamento de pesos (30%) e Early Stopping [2], [16].

Para atingir o objetivo de classificação binária, foram adicionadas as seguintes camadas finais em ordem: (1) Agrupamento por média global 2D (*Global Average Pooling 2D*), (2) uma camada densa com 128 unidades e função de ativação

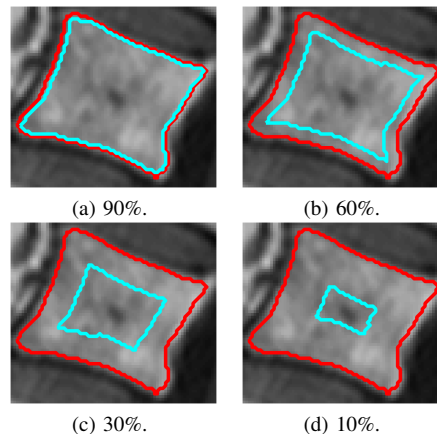


Fig. 1. Recorte da porcentagem da ROI: em vermelho é mostrado o contorno da área original e em ciano da área de recorte obtida.

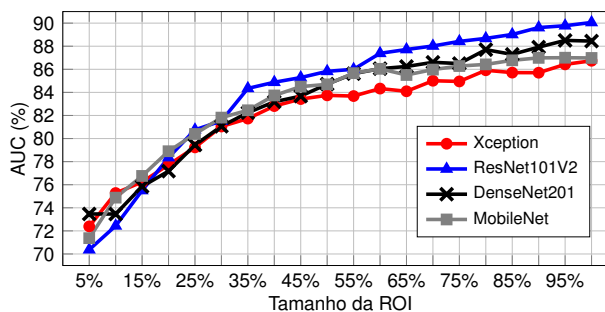


Fig. 2. AUC obtida para cada modelo.

ReLU, (3) uma camada de Dropout com uma taxa de 30% e (4) uma camada densa com duas unidades e função de ativação softmax, adaptado de [2].

Para não enviesar os resultados, todos os conjuntos foram particionados nas mesmas 100 partições sem que existisse intersecção de imagens e de forma que as imagens de um mesmo paciente permanecessem em apenas um conjunto, no entanto por questões de tempo o treinamento foi efetuado até 50 partições, a estratégia de treinamento aplicada foi a de validação-cruzada com 80% das imagens utilizadas para treinamento e 20% para validação dos modelos [2].

### C. Métricas utilizadas

Quanto à validação dos modelos já treinados foram aplicadas as métricas bem estabelecidas para aplicações da área sendo elas: Acurácia, Precisão, Especificidade, Sensibilidade, Medida F (FM) e Área Sob a Curva (do inglês, *Area Under the Curve* - AUC) [17].

## III. RESULTADOS E DISCUSSÕES

Para análise e exibição dos resultados foram abordados apenas quatro modelos de DNN que apresentaram o melhor desempenho em [2], sendo eles: Xception, ResNet101V2, DenseNet201 e MobileNet.

<sup>2</sup><https://keras.io/api/applications>

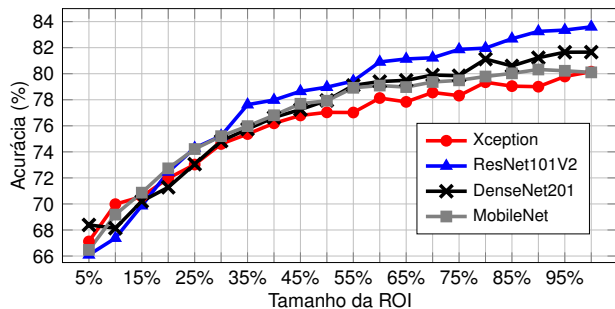


Fig. 3. Acurácia obtida para cada modelo.

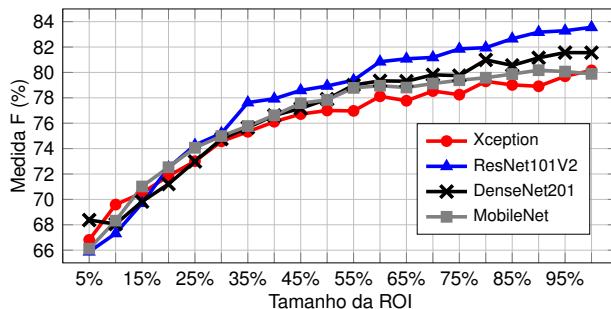


Fig. 7. Medida F obtida para cada modelo.

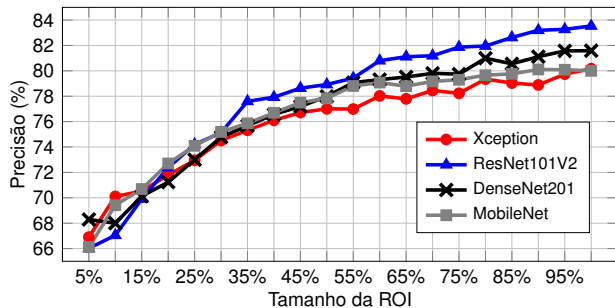


Fig. 4. Precisão obtida para cada modelo.

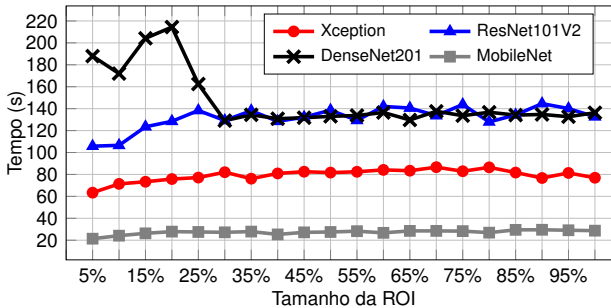


Fig. 8. Tempo médio de treinamento para cada modelo.

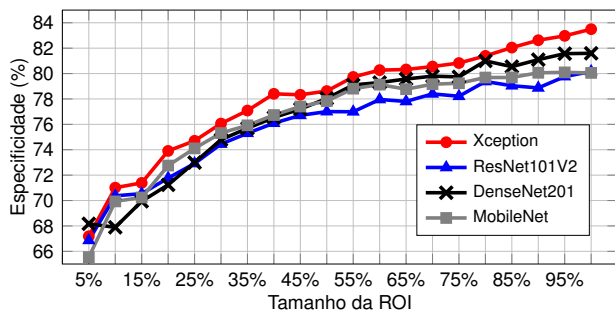


Fig. 5. Especificidade obtida para cada modelo.

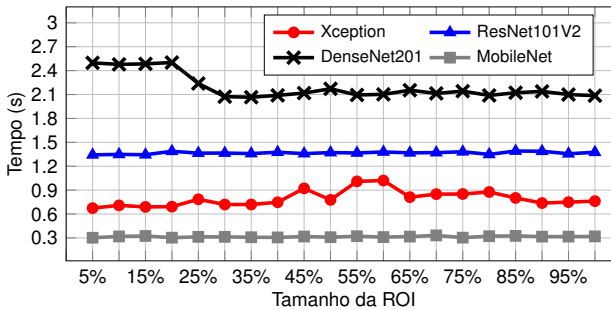


Fig. 9. Tempo médio de validação para cada modelo.

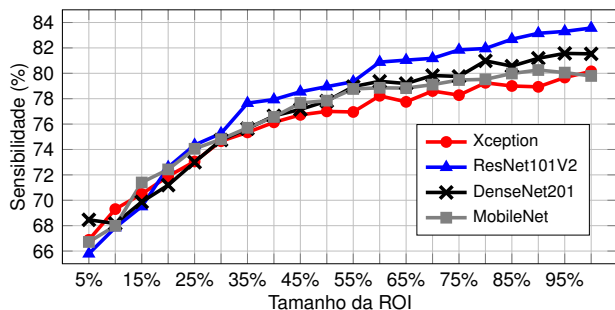


Fig. 6. Sensibilidade obtida para cada modelo.

### A. Comparação geral

As Figuras 2 a 7 exibem os resultados obtidos para cada métrica aplicada durante a validação dos modelos, sendo que o resultado por subconjunto foi obtido a partir da média de todas as métricas em cada partição.

Note que a partir de 60% da ROI o aumento no desempenho passa a ser mais discreto, ou seja, a partir destes pontos o aumento do tamanho dos ROIs pode ter deixado de influenciar no desempenho dos modelos.

As Figuras 8 e 9 exibem respectivamente a média do tempo de treinamento e validação de cada modelo levando em consideração os valores obtidos em cada partição, sendo possível perceber que o tempo ficou relativamente estável mesmo com as alterações de tamanho das ROIs.

É possível notar que apesar de ter obtido um dos piores resultados dentre os modelos analisados, a MobileNet apresentou o menor tempo de execução podendo ser considerado como um bom modelo para aplicação.

### B. Validação estatística

A validação estatística foi realizada utilizando o método de Wilcoxon com uma significância de 99% e adotando uma abordagem bicaudal. Para realização dos testes estatísticos

TABLE I  
VALORES  $p$  OBTIDOS PARA A MENOR PORCENTAGEM SEM DIFERENÇA SIGNIFICANTE EM RELAÇÃO AO TAMANHO DA IMAGEM ORIGINAL (100%)

Modelo	Tamanho mínimo da ROI	Medidas (valores- $p^a$ )					
		Acurácia	Precisão	Sensibilidade	Especificidade	Medida F	AUC
ResNet101V2	85%	✓ 0.060	✓ 0.059	✓ 0.067	✓ 0.063	✓ 0.062	✓ 0.051
DenseNet201	80%	✓ 0.179	✓ 0.126	✓ 0.182	✓ 0.122	✓ 0.153	✓ 0.039
MobileNet	55%	✓ 0.046	✓ 0.039	✓ 0.063	✓ 0.030	✓ 0.058	✓ 0.015
Xception	80%	✓ 0.270	✓ 0.273	✓ 0.217	✓ 0.311	✓ 0.251	✓ 0.196

**Null hypothesis  $H_0$**

“ $a$  e  $b$  são amostras de distribuição contínua e possuem medianas iguais.”

<sup>a</sup>✓ indica não rejeição de  $H_0$  em um intervalo de confiança de 99%.

foram utilizados os dados de validação de cada modelo para as porcentagens inferiores à original (5% a 95%) em comparação com a imagem original (100%).

Sendo possível definir um tamanho mínimo para as ROIs já que os testes estatísticos não conseguiram identificar uma diferença significativa entre o exposto na Tabela I e os resultados das imagens originais.

#### IV. CONCLUSÃO

Em suma, os modelos estudados apresentaram resultados experimentais próximos aos obtidos por Ramos et al. (2022) [2], além de desempenhar a classificação das imagens em intervalos de tempo pequenos e estáveis compactuando com a emergência e volatilidade de um ambiente médico. Por se tratar de uma reprodução do trabalho de Ramos et al. (2022) [2], apenas com variações das ROIs, os resultados obtidos para as imagens originais não foi reportado.

A validação estatística corrobora com a definição de um tamanho mínimo para as ROIs sem que existam diferenças significativas em comparação com o uso das imagens originais. Para facilitar a replicação do processo foi criado um repositório no site GitHub<sup>3</sup> com os códigos utilizados, parte da base de imagens e os resultados obtidos.

#### REFERENCES

[1] E. López, E. Ibarz, A. Herrera, S. Puértolas, S. Gabarre, Y. Más, J. Mateo, J. Gil-Albarova, and L. Gracia, “A predictive mechanical model for evaluating vertebral fracture probability in lumbar spine under different osteoporotic drug therapies,” *Computer methods and programs in biomedicine*, vol. 131, pp. 37–50, 2016.

[2] J. S. Ramos, E. J. de Aguiar, I. V. Belizario, M. V. Costa, J. G. Maciel, M. T. Cazzolato, C. Traina, M. H. Nogueira-Barbosa, and A. J. Traina, “Analysis of vertebrae without fracture on spine mri to assess bone fragility: A comparison of traditional machine learning and deep learning,” in *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2022, pp. 78–83.

[3] K. Hammernik, T. Ebner, D. Stern, M. Urschler, and T. Pock, “Vertebrae segmentation in 3d ct images based on a variational framework,” in *Recent advances in computational methods and clinical applications for spine imaging*. Springer, 2015, pp. 227–233.

[4] H.-Z. Wu, X.-F. Zhang, S.-M. Han, L. Cao, J.-X. Wen, W.-J. Wu, and B.-L. Gao, “Correlation of bone mineral density with mri t2\* values in quantitative analysis of lumbar osteoporosis,” *Archives of Osteoporosis*, vol. 15, no. 1, pp. 1–7, 2020.

[5] N. Binkley, S. Morin, P. Martineau, L. Lix, D. Hans, and W. Leslie, “Frequency of normal bone measurement in postmenopausal women with fracture: a registry-based cohort study,” *Osteoporosis International*, vol. 31, no. 12, pp. 2337–2344, 2020.

[6] M. Tellache, M. Pithioux, P. Chabrand, and C. Hochard, “Femoral neck fracture prediction by anisotropic yield criteria,” *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique*, vol. 18, no. 1, pp. 33–41, 2009.

[7] J. G. Maciel, I. M. d. Araújo, L. C. Trazzi, P. M. d. Azevedo-Marques, C. E. G. Salmon, F. J. A. d. Paula, and M. H. Nogueira-Barbosa, “Association of bone mineral density with bone texture attributes extracted using routine magnetic resonance imaging,” *Clinics*, vol. 75, 2020.

[8] L. He, Z. Liu, C. Liu, Z. Gao, Q. Ren, L. Lei, and J. Ren, “Radiomics based on lumbar spine magnetic resonance imaging to detect osteoporosis,” *Academic radiology*, vol. 28, no. 6, pp. e165–e171, 2021.

[9] A. Y. Kim, M. A. Yoon, S. J. Ham, Y. C. Cho, Y. Ko, B. Park, S. Kim, E. Lee, R. W. Lee, C. G. Chee *et al.*, “Prediction of the acuity of vertebral compression fractures on ct using radiologic and radiomic features,” *Academic Radiology*, 2022.

[10] J. S. Ramos, J. G. Maciel, M. T. Cazzolato, C. Traina, M. H. Nogueira-Barbosa, and A. J. Traina, “Beaut: a radiomic approach to identify potential lumbar fractures in magnetic resonance imaging,” in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2021, pp. 389–394.

[11] J. d. S. Ramos, J. G. Maciel, A. J. M. Traina, P. M. d. Azevedo-Marques, and M. H. Nogueira-Barbosa, “Spine mri texture analysis and prediction of osteoporotic vertebral fracture,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, pp. S115–S116, 2021.

[12] J. S. Ramos, C. Y. Watanabe, M. H. Nogueira-Barbosa, and A. J. Traina, “Bgrowth: an efficient approach for the segmentation of vertebral compression fractures in magnetic resonance imaging,” in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, 2019, pp. 220–227.

[13] J. S. Ramos, M. T. Cazzolato, B. S. Faiçal, M. H. Nogueira-Barbosa, C. Traina, and A. J. Traina, “3dbgrowth: volumetric vertebrae segmentation and reconstruction in magnetic resonance imaging,” in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2019, pp. 435–440.

[14] J. S. Ramos, M. T. Cazzolato, B. S. Faiçal, O. A. Linares, M. H. Nogueira-Barbosa, C. Traina, and A. J. Traina, “Fast and smart segmentation of paraspinal muscles in magnetic resonance imaging with cleverseg,” in *2019 32nd SIBGRAP Conference on Graphics, Patterns and Images (SIBGRAP)*. IEEE, 2019, pp. 76–83.

[15] J. S. Ramos, M. T. Cazzolato, M. H. Nogueira-Barbosa, and A. J. Traina, “Fine: improving time and precision of segmentation techniques for vertebral compression fractures in mri,” in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, 2020, pp. 198–201.

[16] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer *et al.*, “Common pitfalls and recommendations for using machine learning to detect and prognosticate for covid-19 using chest radiographs and ct scans,” *Nature Machine Intelligence*, vol. 3, no. 3, pp. 199–217, 2021.

[17] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.

<sup>3</sup><https://github.com/thiago-coucello/pibic-2022-2023>