

Classificação de Graus ISUP em Imagens Histopatológicas de Próstata utilizando EfficientNet e Aprendizado com Perda Ordinal Focal

Woshington V. S. Rodrigues¹, Armando L. Borges²,
Jose D. Araújo², João O. B. Diniz³, Antonio O. C. Filho²

¹ Instituto Federal do Piauí (IFPI) - Picos, PI - Brasil

woshington@ifpi.edu.br

² Universidade Federal do Piauí (UFPI) - Picos, PI - Brasil

{armandoluzborges, jose.denes, antoniooseas}@ufpi.edu.br

³ Instituto Federal do Maranhão (IFMA) - São Luís, MA - Brasil

joao.bandeira@ifma.edu.br

Abstract. *The manual histopathological evaluation of prostate cancer in WSIs is subject to interobserver and intraobserver variability. To mitigate this issue, this work proposes the automatic classification of ISUP grade groups using Convolutional Neural Networks. The methodology employed the PANDA dataset, applying entropy-based noise reduction and patch extraction. EfficientNet models (B0, B3, and B7) were trained using the Ordinal Focal Loss function to preserve the ordinal progression of the disease and mitigate class imbalance. The best-performing model was EfficientNet-B0, which achieved a quadratic kappa of 0.856 and an accuracy of 66.9%. The approach preserved the ordinal integrity of the grades, making extreme errors infrequent and demonstrating its potential clinical viability.*

Resumo. *A avaliação histopatológica manual do câncer de próstata em WSIs está sujeita a variabilidades interobservador e intraobservador. Para mitigar esse problema, este trabalho propõe a classificação automática dos grupos de graus ISUP utilizando Redes Neurais Convolucionais. A metodologia utilizou o dataset PANDA, aplicando a redução de ruído por entropia e a extração de patches. Os modelos EfficientNet (B0, B3 e B7) foram treinados com a função Ordinal Focal Loss para preservar a progressão da doença e mitigar o desbalanceamento das classes. O melhor modelo foi EfficientNet-B0 que alcançou kappa quadrático de 0,856 e acurácia de 66,9%. A abordagem preservou a integridade ordinal dos graus, tornando pouco frequentes os erros extremos e demonstrando sua viabilidade clínica.*

1. Introdução

O câncer de próstata é uma das neoplasias mais comuns em homens - em 2022, foi a quarta neoplasia mais frequente no mundo, representando 7,3% de todos os casos [Bray et al. 2024]. O diagnóstico histopatológico baseia-se no sistema de graduação proposto pela *International Society of Urological Pathology* (ISUP), derivado do escore de

Gleason e organizado em cinco grupos (1 a 5) de progressão tumoral [Epstein et al. 2016]. A avaliação manual de imagens de lâminas inteiras (*Whole-Slide Images* - WSIs) é exaustiva e sujeita a variabilidades inter e intraobservador, o que compromete o manejo clínico [Bulten et al. 2020]. Nesse contexto, as Redes Neurais Convolucionais (RNCs) têm representado o *estado-da-arte* para automatizar a análise histopatológica [Campanella et al. 2019]. Contudo, trabalhos anteriores frequentemente ignoram a natureza ordinal do sistema ISUP, utilizam funções de perda inadequadas para dados desbalanceados e não tratam ruídos inerentes ao processo de rotulação [Xiang et al. 2023, Afifi et al. 2024, Alici-Karaca and Akay 2024].

Este trabalho propõe um pipeline para a classificação automática dos graus ISUP (0-5) em WSIs prostáticas. As principais contribuições são: (i) uma função de perda híbrida que combina penalização ordinal e *Focal Loss*, preservando a hierarquia biológica da doença e mitigando o desbalanceamento de classes; (ii) um método de redução de ruído baseado na entropia de Shannon das predições de um conjunto de modelos EfficientNet; e (iii) uma avaliação comparativa das arquiteturas EfficientNet-B0, B3 e B7.

2. Materiais e Método

Esta seção descreve o pipeline proposto, ilustrado na Figura 1, estruturado em três etapas: (i) materiais, (ii) pré-processamento e (iii) extração de características e treinamento.

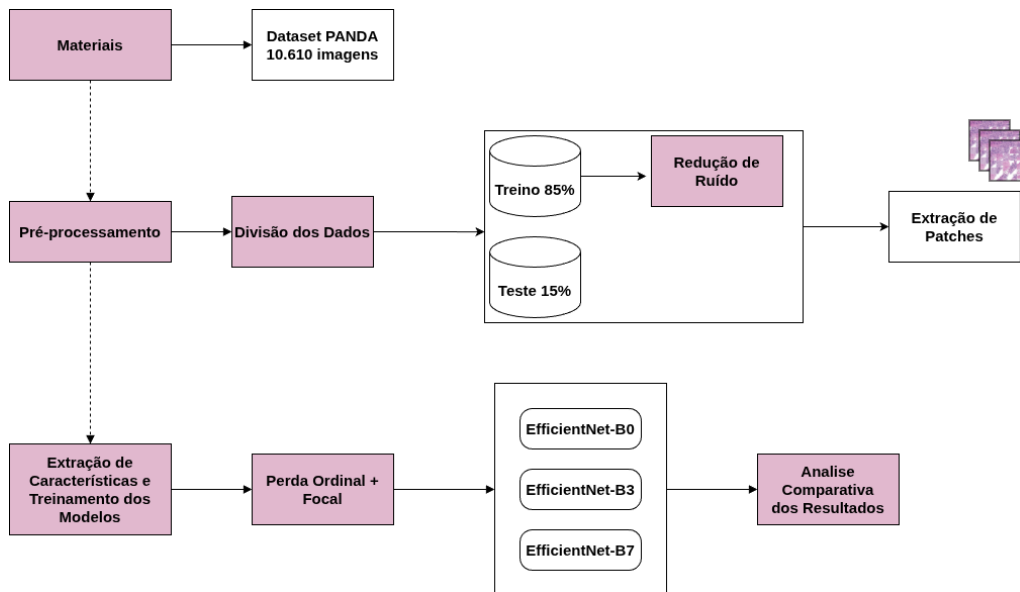


Figura 1. Pipeline proposto.

2.1. Materiais

Foi utilizada a base de dados PANDA (*Prostate Cancer Grade Assessment*), composta por aproximadamente 10.616 WSIs de biópsias prostáticas [PANDA Challenge 2020], coletadas em dois centros médicos - *Radboud University Medical Center* e *Karolinska Institute* - introduzindo variabilidade institucional representativa da prática clínica. As imagens são anotadas com escore e grau de Gleason. O conjunto apresenta distribuição desigual entre as seis classes ISUP (0 a 5), com maior concentração nas classes ISUP 0 e 1, além de ruídos como artefatos visuais e inconsistências de rotulação (Figura 2).

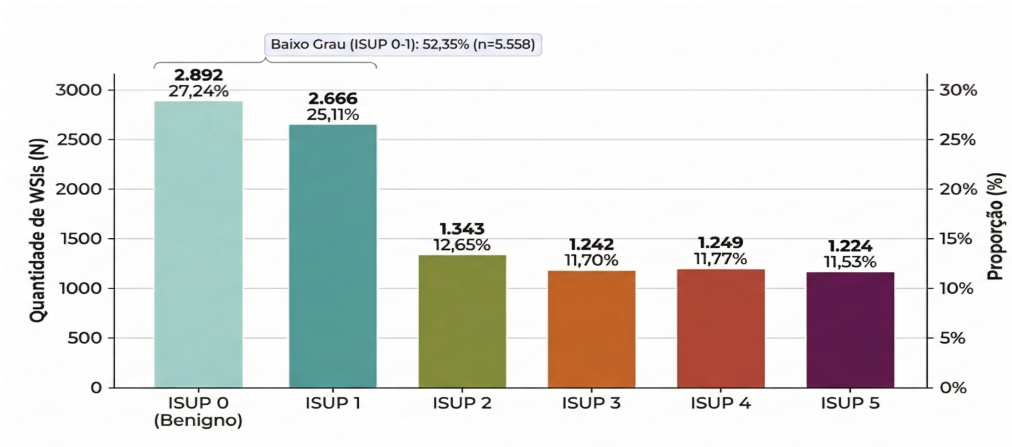


Figura 2. Distribuição das amostras da base PANDA por grau ISUP.

2.2. Pré-processamento

Os dados foram divididos em conjuntos de treinamento e teste. Na sequência, aplicou-se redução de ruído baseada em entropia e, por fim, extração de *patches*.

Redução de ruído por entropia. Quatro modelos EfficientNet (B0-B3) foram usados para calcular a entropia de Shannon e quantificar a incerteza das predições [Ali et al. 2023]. A entropia média dos quatro modelos foi calculada por amostra, e o percentil superior de 10% formou um banco de amostras candidatas a ruído. Avaliaram-se quatro cenários (remoção de 0%, 20%, 50% e 100% dessas amostras), sendo a remoção de 20% (180 amostras) a que produziu melhor desempenho.

Extração de *patches*. Como as WSIs são de alta dimensionalidade e o tecido ocupa apenas parte da lâmina, foram geradas novas imagens por reorganização em *patches*. Avaliaram-se configurações 5×5, 6×6 e 7×7; a configuração 6×6 (36 *patches*, 1536×1536 px) foi selecionada por equilibrar cobertura tecidual, qualidade e eficiência computacional.

2.3. Extração de características e treinamento

A EfficientNet-B0 foi utilizada como modelo-base para ajuste de hiperparâmetros (*fine-tuning* com 150 camadas descongeladas, taxa de aprendizado 3×10^{-4} , *dropout* 0,4, 50 épocas, otimizador Adam). Essa configuração foi então estendida às EfficientNet-B3 e B7.

A função de perda baseline (*Binary Cross-Entropy*) foi substituída por uma formulação híbrida que reflete a natureza ordinal do ISUP. Cada grau $k \in \{0, \dots, 5\}$ é codificado como vetor binário cumulativo (k primeiros elementos iguais a 1), e a perda ordinal é aplicada a cada limiar, impondo penalização proporcional à distância entre classes. Para mitigar o desbalanceamento, a perda ordinal foi combinada à *Focal Loss* [Lin et al. 2017] ($\gamma = 2,0$, $\alpha = 0,25$), priorizando amostras de maior incerteza diagnóstica. O desempenho foi avaliado por **acurácia** e pelo **Kappa Quadrático Ponderado** (κ_{quad}), com intervalos de confiança estimados via bootstrap não paramétrico (1.000 reamostragens, IC de 90%).

3. Resultados

Os experimentos foram conduzidos em uma estação Linux com GPU RTX 3060, CPU AMD Ryzen 5 5600X e 32 GB de RAM DDR4, utilizando PyTorch, CUDA e OpenSlide.

3.1. Impacto da Remoção de Ruído e das Funções de Perda

A Tabela 1 apresenta a evolução incremental do desempenho da EfficientNet-B0. A remoção de ruído elevou o κ_{quad} de 0,826 para 0,833. A substituição da BCE pela perda ordinal promoveu um salto para 0,851, com o menor desvio padrão (0,009), indicando maior estabilidade. Por fim, a estratégia híbrida (Ordinal + Focal) resultou no melhor desempenho: $\kappa_{\text{quad}} = 0,856$ e acurácia de 66,9%, ganho absoluto de 3% em relação ao *baseline*.

Configuração	Kappa Quadrático			Acurácia (%)		
	Resultado	Std	IC 90%	Resultado	Std	IC 90%
BCE (<i>baseline</i>)	0,826	0,012	[0,806; 0,846]	0,592	0,012	[0,572; 0,612]
BCE + Remoção de Ruído	0,833	0,013	[0,812; 0,853]	0,638	0,116	[0,619; 0,657]
Perda Ordinal	0,851	0,009	[0,833; 0,869]	0,608	0,126	[0,587; 0,628]
Ordinal + Focal (Híbrida)	0,856	0,011	[0,833; 0,876]	0,669	0,011	[0,635; 0,683]

Tabela 1. Impacto da função de perda no desempenho da EfficientNet-B0.

3.2. Comparação entre Arquiteturas EfficientNet

A mesma configuração foi aplicada às variantes B3 e B7 (Tabela 2). A EfficientNet-B0 apresentou o melhor desempenho, tanto em κ_{quad} quanto em acurácia. As arquiteturas mais profundas não apresentaram ganho relevante, houve inclusive leve degradação, sugerindo que modelos com maior número de parâmetros requerem ajustes mais refinados ou conjuntos de dados maiores para atingir seu potencial. Assim, a B0 demonstrou o melhor equilíbrio entre desempenho, estabilidade e custo computacional.

Arquitetura	Kappa Quadrático			Acurácia (%)		
	Resultado	Std	IC 90%	Resultado	Std	IC 90%
EfficientNet-B0	0,856	0,011	[0,833; 0,876]	0,669	0,119	[0,635; 0,683]
EfficientNet-B3	0,846	0,010	[0,829; 0,861]	0,659	0,125	[0,554; 0,594]
EfficientNet-B7	0,844	0,011	[0,821; 0,865]	0,602	0,112	[0,578; 0,628]

Tabela 2. Comparação entre variantes da EfficientNet.

Em termos clínicos, os modelos concentraram os erros entre classes adjacentes (especialmente ISUP 2-4), comportamento esperado em problemas de classificação ordinal e que confirma a eficácia da função de perda híbrida na preservação da hierarquia.

3.3. Comparação com o Estado da Arte

A Tabela 3 compara a abordagem proposta com trabalhos recentes. Embora [Xiang et al. 2023] reporte $\kappa = 0,931$ em validação interna, o desempenho cai para 0,801 em validação externa. Nossa abordagem alcança $\kappa = 0,856$ diretamente nos dados públicos ruidosos do PANDA, indicando maior robustez prática. O modelo proposto

realiza a graduação completa dos seis níveis ISUP (0–5) e incorpora explicitamente a natureza ordinal da doença, diferenciando-se das abordagens binárias ou que utilizam funções de perda inadequadas.

Trabalho	Base	Modelo	Tarefa	Acurácia	Kappa
[Xiang et al. 2023]	PANDA	ResNet50 + GCN	ISUP (6 níveis)	—	0,931 (Int.) / 0,801 (Ext.)
[Alici-Karaca and Akay 2024]	DiagSet	EfficientNet-B4 + ECA	Binária / 4 classes	0,961 / 0,948	—
[Kosoko et al. 2024]	SICAPv2	VGG19	Escore de Gleason	0,780	—
[Afifi et al. 2024]	PANDA	InceptionResNetV2	Escore de Gleason	0,918	0,608
Proposta	PANDA	EfficientNet-B0 + Ordinal Focal	ISUP (6 níveis)	0,669	0,856

Tabela 3. Comparação com trabalhos da literatura.

4. Conclusão

Este trabalho apresentou uma metodologia para a classificação automática dos graus ISUP em imagens histopatológicas de próstata, abordando desafios como desbalanceamento de classes, ruído nos rótulos e a natureza ordinal da progressão tumoral. A função de perda híbrida (Ordinal + Focal) foi o principal fator de melhoria, elevando o κ_{quad} de 0,826 para 0,856 (+3%) e a acurácia de 0,592 para 0,669 (+7%) em relação ao *baseline*. Destaca-se que a perda ordinal isolada, combinada à remoção de ruído, já promoveu ganho de 2,5% no κ_{quad} com alta estabilidade (desvio padrão de 0,009). A EfficientNet-B0 demonstrou o melhor equilíbrio entre desempenho, estabilidade e custo computacional, com os erros concentrados entre classes adjacentes — comportamento coerente com a gradação ordinal do ISUP. Como trabalhos futuros, propõe-se investigar arquiteturas baseadas em *Vision Transformers* e estratégias de *ensemble* para aprimorar a classificação dos graus intermediários e ampliar a generalização clínica do modelo.

Referências

- Afifi, A., Alrahmawy, M., and El-Baz, A. (2024). Prostate cancer detection using deep learning models on histopathological image slides: An experimental analysis. *International Journal of Advanced Computer Science and Applications*, 15(1):xxx–xxx.
- Albahri, O. S., Albahri, A. S., Mohammed, K. I., Zaidan, A., Zaidan, B. B., Hashim, M., Salman, O. H., Alaa, M., and Alsalem, M. A. (2022). Expert system research for prostate cancer: A literature review. *Computers in Biology and Medicine*, 145:105487.
- Ali, I., Khan, A., Khan, M., Ahmad, R., and Ullah, I. (2023). Shannon entropy in artificial intelligence and its applications based on information theory. *Entropy*, 25(2):364.
- Alici-Karaca, D. and Akay, B. (2024). An efficient deep learning model for prostate cancer diagnosis. *IEEE Access*, 12:150776–150790.
- Araújo, F. H., Silva, R. R., Medeiros, F. N., Neto, J. F. R., Oliveira, P. H. C., Bianchi, A. G. C., and Ushizima, D. (2021). Active contours for overlapping cervical cell segmentation. *International Journal of Biomedical Engineering and Technology*, 35(1):70–92.
- Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A. F., and Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, 16(5):412–424.

- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., and Jemal, A. (2024). Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3):229–263.
- Bulten, W., Pinckaers, H., van Boven, H., Vink, R., de Bel, T., van Ginneken, B., van der Laak, J., Hulsbergen-van de Kaa, C., and Litjens, G. (2020). Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *The Lancet Oncology*, 21(2):233–241.
- Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Silva, V. W. K., Busam, K. J., Brogi, E., Reuter, V. E., Klimstra, D. S., and Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, 25(9):1301–1309.
- Chen, X. (2026). A transfer learning-based deep focal multiclass network for psychological emotion recognition in community-correction populations. *Alexandria Engineering Journal*, 135:235–242.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4):213–220.
- Dina, A. S., Siddique, A., and Manivannan, D. (2023). A deep learning approach for intrusion detection in internet of things using focal loss function. *Internet of Things*, 22:100699.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.
- Epstein, J. I., Egevad, L., Amin, M. B., Delahunt, B., Srigley, J. R., and Humphrey, P. A. (2016). The 2014 International Society of Urological Pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma: Definition of grading patterns and proposal for a new grading system. *The American Journal of Surgical Pathology*, 40(2):244–252.
- Kosoko, I., Garg, A., Jain, S., and Hewage, P. (2024). Leveraging deep learning and explainable ai for diagnosis of prostate cancer. In *Proceedings of the International Conference on Applied Artificial Intelligence in Medical Imaging*, Bolton, United Kingdom.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988.
- PANDA Challenge (2020). Prostate cANcer grade assessment challenge. <https://www.kaggle.com/c/prostate-cancer-grade-assessment>. Accessed: 2024-01-01.
- Siegel, R. L., Giaquinto, A. N., and Jemal, A. (2024). Cancer statistics, 2024. *CA: A Cancer Journal for Clinicians*, 74(1):12–49.
- Xiang, J., Wang, X., Wang, X., Zhang, J., Yang, S., Yang, W., Han, X., and Liu, Y. (2023). Automatic diagnosis and grading of prostate cancer with weakly supervised learning on whole slide images. *Computers in Biology and Medicine*, 152:106340.