

Impacto do Aumento de Dados por Ruído Artificial para Classificação de Biópsias de Câncer de Estômago

Quézia P. Silva¹, Alexei C. Machado^{1,2}

¹Departamento de Ciência da Computação
Pontifícia Universidade Católica de Minas Gerais (PUC Minas)
30.535-901 – Belo Horizonte – MG – Brasil

²Departamento de Anatomia e Imagem – Universidade Federal de Minas Gerais
Av Alfredo Balena 190 – 30130-100 – Belo Horizonte – MG – Brasil

qsilva@sga.pucminas.br, alexei@sga.pucminas.br

Abstract. *Gastric cancer is among the leading causes of death worldwide. However, the application of deep learning models to support its diagnosis is often limited by the scarcity and imbalance of labeled data. This work investigates the impact of data augmentation techniques based on artificial noise on the classification of histopathological images of gastric biopsies. Three deep neural network architectures (InceptionV3, VGG16, and ResNet50) were evaluated in scenarios with and without data augmentation. In addition, the effect of using different types and intensities of noise in isolation was analyzed. The results demonstrate that the combined use of data augmentation techniques significantly improves the performance of the models, although the isolated use of artificial noise results in a drop in performance.*

Resumo. *O câncer gástrico está entre as principais causas de morte por câncer no mundo. No entanto, a aplicação de modelos de aprendizado profundo para seu diagnóstico é frequentemente limitada pela escassez e desbalanceamento de dados rotulados. Este trabalho investiga o impacto de técnicas de aumento de dados baseadas em ruído artificial na classificação de imagens histopatológicas de biópsias gástricas. Foram avaliadas três arquiteturas de redes neurais profundas (InceptionV3, VGG16 e ResNet50) em cenários com e sem aumento de dados. Os resultados demonstram que o uso combinado de técnicas de aumento de dados melhora significativamente o desempenho dos modelos, embora o uso isolado de ruído artificial resulte em queda de desempenho.*

1. Introdução

O câncer gástrico é considerado uma das principais causas de incidência e mortalidade por câncer em todo o mundo [Wang et al. 2023]. Apesar dos avanços na área médica, o diagnóstico precoce ainda representa um desafio, especialmente na análise de imagens histopatológicas, que exige alto nível de especialização.

Nos últimos anos, técnicas de aprendizado profundo têm sido amplamente aplicadas na classificação de imagens médicas, apresentando resultados promissores [Rai et al. 2024]. No entanto, o desempenho desses modelos depende diretamente da disponibilidade de grandes volumes de dados rotulados, o que nem sempre é viável em contextos médicos [Loddo et al. 2024].

Nesse cenário, técnicas de *data augmentation* têm sido utilizadas para aumentar artificialmente a diversidade dos dados de treinamento, contribuindo para a redução do *overfitting* e melhoria da generalização dos modelos [Shorten and Khoshgoftaar 2019]. Dentre essas técnicas, a adição de ruído artificial tem sido explorada como forma de simular variações presentes em cenários reais [Yang et al. 2021].

Embora estudos indiquem que o uso de ruído pode melhorar o desempenho em determinadas aplicações [Zhao et al. 2024], ainda não está claro em que medida sua aplicação isolada contribui efetivamente para a generalização dos modelos em imagens histopatológicas.

Apesar do uso de bases de dados com número considerável de imagens, como o *GasHisSDB*, a limitação de dados em aplicações médicas não está apenas relacionada à quantidade de amostras, mas principalmente à disponibilidade de dados rotulados de alta qualidade. Esse processo depende da atuação de especialistas, como patologistas, tornando a obtenção de anotações confiáveis um processo custoso e demorado [Loddo et al. 2024].

Diante disso, este trabalho tem como objetivo investigar o impacto do uso de ruído artificial como técnica de aumento de dados na classificação de imagens de biópsias de câncer gástrico. Para isso, são avaliados dois cenários distintos: (i) uso combinado de múltiplas técnicas de aumento de dados e (ii) uso isolado de diferentes tipos e intensidades de ruído.

Como contribuições deste trabalho, destacam-se:

- Avaliação comparativa de três arquiteturas (*InceptionV3*, *VGG16* e *ResNet50*) com e sem aumento de dados;
- Investigação do impacto do uso isolado de diferentes tipos e intensidades de ruído;
- Evidência experimental de que o uso isolado de ruído pode degradar o desempenho dos modelos.

2. Materiais e Métodos

O conjunto de dados utilizado nos experimentos provém do *dataset GasHisSDB* [Hu et al. 2022], composto por 13.124 imagens com evidência de câncer gástrico e 20.160 imagens sem evidência de câncer gástrico com resolução de 160×160 pixels. A Fig. 1 exemplifica duas imagens do *dataset*.

O fluxograma da metodologia proposta é apresentado na Fig. 2, que ilustra os três caminhos adotados para o treinamento dos classificadores: sem aumento de dados; com aumento para balanceamento com ruído; e com aumento combinando ruído e rotações. O processo inicia-se no pré-processamento das imagens, seguido pela preparação dos conjuntos de dados, treinamento com as arquiteturas *InceptionV3*, *VGG16* e *ResNet50*, e finaliza com a avaliação e análise dos resultados.

As distribuições de ruído analisadas foram a Gaussiana (desvio padrão $\sigma = 25$), Sal-e-pimenta (probabilidade = 0,1) e Uniforme no intervalo $[-50, 50]$, combinadas ou não com rotações de 10, 30 e 90 graus. A Fig. 3 demonstra exemplos de aplicação dos ruídos. O conjunto foi dividido em 70% treino, 15% validação e 15% teste. As métricas utilizadas para avaliação de desempenho do classificador foram a acurácia, precisão, *recall* e *F1-score*, através de validação cruzada com 10 dobras .

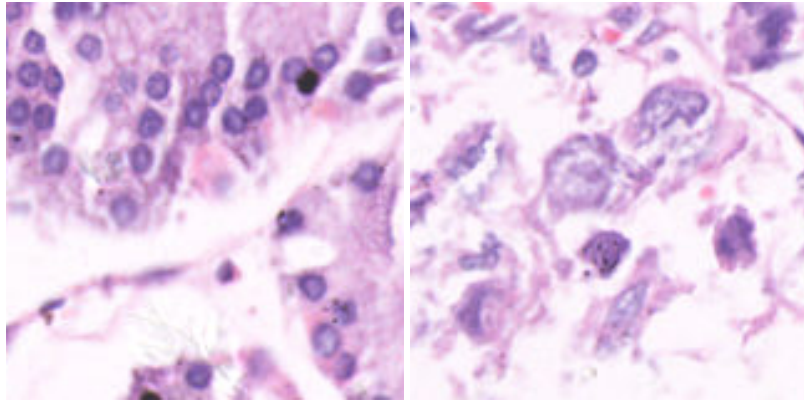


Figura 1. Exemplos de imagens do dataset sem câncer gástrico (esquerda) e com câncer gástrico (direita).

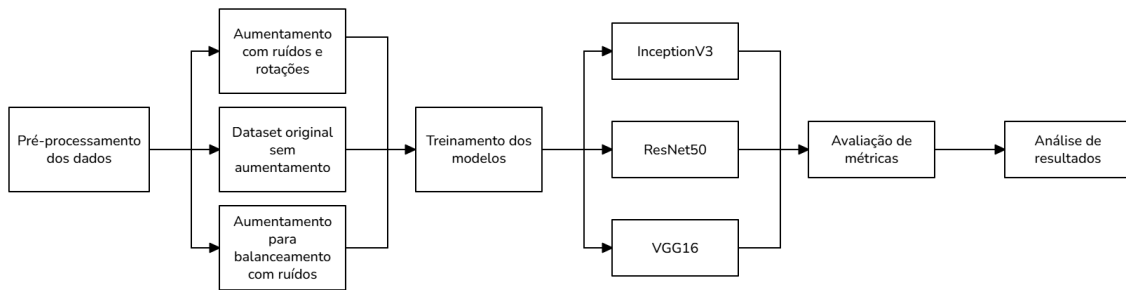


Figura 2. Fluxograma da metodologia proposta, ilustrando os três caminhos adotados para o treinamento dos classificadores: sem aumento de dados, com aumento para balanceamento via ruído, e com aumento combinando ruído e rotações. O processo inicia-se no pré-processamento das imagens, seguido pela preparação dos conjuntos de dados, treinamento com as arquiteturas *InceptionV3*, *VGG16* e *ResNet50*, e finaliza com a avaliação e análise dos resultados.

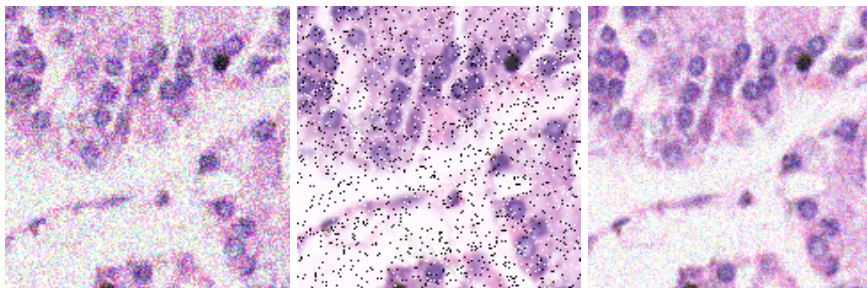


Figura 3. Exemplo de aplicação de diferentes tipos de ruído: uniforme (esquerda), sal-e-pimenta (centro) e Gaussiano (direita).

As arquiteturas pré-treinadas *InceptionV3*, *VGG16* e *ResNet50* foram utilizadas como base para os experimentos. Inicialmente, apenas as camadas posteriores foram treinadas, enquanto as camadas convolucionais foram mantidas congeladas a fim de que os pesos previamente aprendidos com o *dataset ImageNet* fossem preservados. Posteriormente, foi realizado *fine-tuning* por meio do descongelamento das últimas camadas convolucionais.

Os modelos foram treinados com otimizador Adam, função de perda *binary cross*-

entropy, até 100 épocas, utilizando *early stopping*. O processo de treinamento de todos os modelos foi acompanhado por meio da técnica de *Early Stopping* com monitoramento da função de perda, a fim de que o *overfitting* fosse evitado. O treinamento era interrompido automaticamente quando nenhuma melhoria na função de perda da validação fosse observada por 10 épocas consecutivas. Para a avaliação dos modelos, foi utilizada validação cruzada com 10 *folds* sobre o conjunto de treinamento. Em cada iteração, os dados foram divididos em subconjuntos de treino e validação, mantendo-se a proporção entre as classes. O conjunto de teste (15%) foi mantido separado e utilizado exclusivamente para avaliação final, garantindo que não houvesse vazamento de dados durante o treinamento.

3. Resultados Experimentais

Foram realizados experimentos em três cenários: sem aplicação de técnicas de aumento de dados; aumento de dados através de rotações e aplicação de ruído; e apenas com aplicação de ruído.

3.1. Experimentos com aumento por rotação e aplicação de ruído

A Tabela 1 apresenta as métricas obtidas no conjunto de teste para as três arquiteturas avaliadas. O melhor desempenho foi obtido pela *ResNet50* com aumento de dados por rotações e adição de ruídos, atingindo 0,91 de acurácia. Sem aumento, observou-se queda significativa de desempenho, evidenciando o impacto positivo do aumento.

Modelo	Aumento	Acurácia	Precisão	Recall	F1-score
InceptionV3	Sim	0.87	0.84	0.82	0.83
InceptionV3	Não	0.87	0.82	0.80	0.81
VGG16	Sim	0.84	0.82	0.82	0.82
VGG16	Não	0.86	0.76	0.82	0.79
ResNet50	Sim	0.91	0.90	0.85	0.88
ResNet50	Não	0.75	0.69	0.66	0.67

Tabela 1. Desempenho dos classificadores com e sem aumento de dados.

3.2. Experimentos com aumento por aplicação de ruído isolado

Neste conjunto de experimentos, foi avaliado o impacto do aumento de dados por adição isolada de ruído de diferentes magnitudes, a fim de se analisar o comportamento da *Resnet50* na classificação das biópsias. Todos os resultados apresentados na Tabela 2 referem-se ao desempenho no conjunto de teste, obtido após o treinamento final de cada configuração com validação cruzada.

De modo geral, os valores obtidos foram inferiores aos alcançados nos experimentos sem qualquer aumento de dados, com *F1-score* variando entre 0,43 e 0,44. Os testes *t* bilaterais entre os resultados com e sem aplicação de ruído indicaram p-valores inferiores a 0,001 em todos os casos, evidenciando que a introdução isolada de ruído comprometeu significativamente o desempenho em comparação ao cenário base, sem qualquer técnica de aumento de dados. Alguns dos possíveis fatores que contribuíram para esse desempenho inferior incluem:

Ruído	Valor	Recall	Precisão	F1-score	Acurácia.	P-valor
Gaussiano	5	0.47 (0.02)	0.40 (0.02)	0.44 (0.02)	0.52 (0.02)	< 0.001
	10	0.49 (0.02)	0.40 (0.02)	0.44 (0.02)	0.51 (0.02)	< 0.001
	15	0.49 (0.03)	0.40 (0.02)	0.44 (0.03)	0.51 (0.02)	< 0.001
	25	0.49 (0.03)	0.40 (0.03)	0.44 (0.03)	0.51 (0.03)	< 0.001
	50	0.49 (0.02)	0.40 (0.02)	0.44 (0.02)	0.52 (0.02)	< 0.001
<i>Sal-e-Pimenta</i>	10%	0.46 (0.02)	0.39 (0.02)	0.43 (0.02)	0.51 (0.02)	< 0.001
	30%	0.47 (0.02)	0.40 (0.02)	0.43 (0.02)	0.52 (0.02)	< 0.001
	50%	0.46 (0.02)	0.40 (0.02)	0.43 (0.02)	0.51 (0.02)	< 0.001
	70%	0.46 (0.03)	0.40 (0.03)	0.43 (0.03)	0.52 (0.03)	< 0.001
	85%	0.47 (0.02)	0.40 (0.02)	0.43 (0.02)	0.52 (0.02)	< 0.001
Uniforme	[-5,5]	0.45 (0.02)	0.40 (0.02)	0.42 (0.02)	0.52 (0.02)	< 0.001
	[-25,25]	0.50 (0.03)	0.40 (0.02)	0.44 (0.03)	0.51 (0.02)	< 0.001
	[-45,45]	0.49 (0.03)	0.40 (0.02)	0.44 (0.02)	0.51 (0.02)	< 0.001
	[-50,50]	0.50 (0.03)	0.40 (0.02)	0.44 (0.03)	0.51 (0.02)	< 0.001
	[-75,75]	0.49 (0.03)	0.40 (0.02)	0.44 (0.03)	0.51 (0.02)	< 0.001
Sem ruído	-	0.66 (0.0005)	0.69 (0.37)	0.67 (0.0010)	0.75 (0.0003)	-

Tabela 2. Média(desvio-padrão) para o desempenho da ResNet50 com diferentes tipos e intensidades de ruído. Os p-valores indicam a significância para o teste t bilateral entre s resultados da validação cruzada com e sem aumento.

- 1. Volume reduzido de imagens geradas:** Em cada experimento, apenas uma cópia da imagem original foi modificada com um tipo específico de ruído, resultando em um aumento limitado do volume de dados.
- 2. Objetivo distinto dos testes:** Diferentemente dos testes principais, que buscavam melhorar o desempenho do modelo, o propósito dos experimentos com ruído isolado foi investigar os impactos de diferentes níveis de adição de ruído, ou seja, quais níveis seriam benéficos e até que ponto o modelo se beneficiaria dessa modalidade de aumento de dados.
- 3. Confusão da rede:** Ao introduzir ruído nas imagens, o modelo pode ter considerado esse aspecto de textura como um elemento confundidor para a classificação, enquanto que aumentos por rotação não interferiram no padrão textural das classes.

4. Discussão

Os resultados indicam que a *ResNet50* apresentou melhor desempenho geral, possivelmente devido à sua capacidade de generalização proporcionada pelas conexões residuais. A aplicação combinada de técnicas de aumento contribuiu para reduzir o *overfitting*, enquanto o uso isolado de ruído não se mostrou eficaz, atuando como fator confundidor ao introduzir variações irrelevantes nos dados. Observou-se ainda que rotações apresentam menor custo computacional, com treinamento mais rápido e desempenho semelhante ao obtido com combinação de ruído artificial e rotações. Esses resultados estão alinhados com a literatura, que aponta que o uso combinado de técnicas de aumento de dados melhora a generalização em cenários com dados limitados [Shorten and Khoshgoftaar 2019].

5. Limitações

Como limitações, destaca-se que o aumento de dados com ruído isolado foi realizado apenas até o balanceamento das classes, sem geração de múltiplas variações por imagem, o que pode ter limitado a diversidade do conjunto de treino. Além disso, os experimentos foram conduzidos apenas na sub-base de 160×160 pixels do *GasHisSDB*, restringindo a generalização dos resultados. Por fim, limitações computacionais impediram a avaliação de arquiteturas mais profundas e combinações adicionais de técnicas.

6. Conclusão

Neste trabalho, investigou-se o impacto de técnicas de aumento de dados na classificação de imagens histopatológicas de câncer gástrico utilizando uma sub-base do conjunto *GasHisSDB*. Foram aplicadas rotações e ruído para o treinamento das arquiteturas *InceptionV3*, *VGG16* e *ResNet50*. Os resultados demonstraram que o aumento de dados por rotação contribuiu para a redução do *overfitting* e melhoria do desempenho, destacando-se a *ResNet50*, que atingiu acurácia de até 0,91 com as técnicas aplicadas, frente a 0,75 sem aumento. Nos testes com aplicação isolada de ruídos artificiais, no entanto, houve uma queda de desempenho, indicando que a aplicação isolada de ruído não foi suficiente para melhorar a generalização dos modelos.

Agradecimentos — Alexei M. C. Machado agradece o apoio financeiro do CNPq e da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) projetos APQ-02753-24 e APQ-06556-24.

Referências

- Hu, W., Li, C., Li, X., Rahaman, M. M., Ma, J., Zhang, Y., Chen, H., Liu, W., Sun, C., Yao, Y., Sun, H., and Grzegorzec, M. (2022). Gashissdb: A new gastric histopathology image dataset for computer aided diagnosis of gastric cancer. *Computers in Biology and Medicine*, 142:105207.
- Loddo, A., Usai, M., and Di Ruberto, C. (2024). Gastric cancer image classification: A comparative analysis and feature fusion strategies. *Journal of Imaging*, 10(8).
- Rai, H. M., Yoo, J., Atif Moqurrab, S., and Dashkevych, S. (2024). Advancements in traditional machine learning techniques for detection and diagnosis of fatal cancer types: Comprehensive review of biomedical imaging datasets. *Measurement*, 225:114059.
- Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48.
- Wang, Z., Liu, Y., and Niu, X. (2023). Application of artificial intelligence for improving early detection and prediction of therapeutic outcomes for gastric cancer in the era of precision oncology. *Seminars in Cancer Biology*, 93:83–96.
- Yang, Y., Zhai, G., Hu, M., and Jiang, W. (2021). Adaptive noise injection for quality improvement of blurry images. In Zhai, G., Zhou, J., Yang, H., An, P., and Yang, X., editors, *Digital TV and Wireless Multimedia Communication*, pages 96–106, Singapore. Springer Singapore.
- Zhao, J., Lui, H., Kalia, S., Lee, T. K., and Zeng, H. (2024). Improving skin cancer detection by raman spectroscopy using convolutional neural networks and data augmentation. *Frontiers in Oncology*, 14:1320220. PMID: 38962264.