Deep Learning for Lung Cancer Histopathological Image Classification: Exploring Data Augmentation and Ensemble Learning Strategies

Joyce M. Silva[1], Mateus A. Veloso[2], José D. L. Araújo [3], Antonio O. de Carvalho Filho[1,2,3]

[1]Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal do Piauí - Teresina, Brasil

[2]Programa de Pós-Graduação em Ciência da Computação
Universidade Federal do Piauí - Teresina, Brasil

[3]Departamento de Sistemas de Informação
Universidade Federal do Piauí - Picos, Brasil

{joyce.moura,mateusveloso,jose.denes,antoniooseas}@ufpi.edu.br

Abstract. *Lung cancer is among the leading causes of cancer-related mortality worldwide. In this context, the analysis of histopathological images plays an essential role in diagnosis. In this work, deep learning techniques were investigated for the classification of lung cancer histopathological images, combining data augmentation and ensemble learning strategies. The results demonstrated that these approaches significantly improved model performance, achieving accuracies of up to 100% on images at 20× magnification and 98% on images at 40× magnification. These findings highlight the potential of such techniques as support tools for histopathological analysis.*

Resumo. *O câncer de pulmão está entre as principais causas de mortalidade. Nesse contexto, a análise de imagens histopatológicas desempenha papel essencial no diagnóstico. Neste trabalho, foram investigadas técnicas de deep learning para a classificação de imagens histopatológicas de câncer de pulmão, combinando estratégias de aumento de dados e ensemble learning. Os resultados demonstraram que essas abordagens contribuíram significativamente para o desempenho dos modelos, alcançando acurácias de até 100% em imagens com ampliação de 20x e 98% em imagens de 40x. Os achados evidenciam o potencial dessas técnicas como ferramentas de apoio à análise histopatológica.*

1. Introdução

O câncer de pulmão é uma das principais causas de mortalidade em todo o mundo. De acordo com a Agência Internacional de Pesquisa sobre o Câncer (IARC), o câncer de pulmão foi a neoplasia maligna mais comum globalmente em 2022, com 2,5 milhões de novos casos e 1,8 milhão de óbitos, representando 12,4% de todos os novos diagnósticos [Iarc 2022]. Em escala global, ocupa a primeira posição em incidência entre homens e a terceira entre mulheres [Inca 2022].

O câncer de pulmão pode ser classificado em dois tipos principais: câncer de pulmão de pequenas células (CPCP) e câncer de pulmão de células não pequenas

(CPNPC). Este último corresponde a mais de 85% dos casos e inclui os subtipos adenocarcinoma, carcinoma escamoso, carcinoma de grandes células, entre outros subtipos menos comuns [Acs 2024]. Entre os CPNPCs, o adenocarcinoma pulmonar é o subtipo mais prevalente [Relli et al. 2019]. Embora sua principal causa esteja associada ao tabagismo, cerca de 15% a 20% dos casos ocorrem em não fumantes, sendo geralmente atribuídos a uma combinação de fatores genéticos e ambientais [Myers and Wallen 2023].

Para o diagnóstico, as imagens histopatológicas obtidas por microscopia de amostras de tecido desempenham um papel fundamental na pesquisa médica sobre o câncer. Essas imagens fornecem informações detalhadas sobre a morfologia celular e a organização dos tecidos. Adicionamente, as lâminas preservam a arquitetura tecidual, permitindo uma visualização abrangente das alterações causadas pela doença. Dessa forma, características relevantes, como infiltração linfocítica em tumores, podem ser identificadas diretamente nessas imagens. Apesar de seu grande valor diagnóstico, a análise manual dessas lâminas é desafiadora, demorada e suscetível a erros humanos [Gurcan et al. 2009].

As arquiteturas baseadas em *deep learning* (DL), como *convolutional neural networks* (CNNs) e *vision transformers* (ViTs), têm se destacado no diagnóstico automatizado de imagens médicas, devido à sua capacidade de capturar padrões sutis e complexos em imagens médicas [Takahashi et al. 2024]. Em particular, as CNNs são capazes de aprender representações hierárquicas diretamente dos dados, extraindo características em múltiplos níveis de abstração [Litjens et al. 2017]. Já os ViTs modelam relações globais entre diferentes regiões da imagem por meio do mecanismo de *self-attention*, ampliando a capacidade de representação espacial e a integração de informações contextuais [Dosovitskiy et al. 2020]. Essas características tornam tais arquiteturas particularmente adequadas para a análise de imagens histopatológicas, nas quais padrões locais e globais desempenham papel fundamental na identificação de alterações teciduais.

O presente trabalho investiga diferentes arquiteturas de DL para a classificação de imagens histopatológicas de câncer de pulmão. Para isso, são combinadas estratégias de pré-processamento, aumento de dados e otimização de hiperparâmetros, além da utilização de técnicas de *ensemble learning*, visando melhorar o desempenho e avaliar sua capacidade de auxiliar no diagnóstico automatizado da doença.

2. Trabalhos relacionados

A aplicação de estratégias de DL na análise de imagens histopatológicas tem proporcionado avanços significativos no diagnóstico assistido por computador. Esta seção apresenta estudos recentes que utilizam o mesmo conjunto de dados adotado neste trabalho, o que permite comparações mais consistentes entre diferentes metodologias.

O artigo de [Diosdado et al. 2024] apresenta um conjunto de imagens histopatológicas de câncer de pulmão, obtidas em ampliações de 20x e 40x. Para avaliar o *dataset*, foram utilizados modelos baseados em *ResNet-50* e Aprendizado por Múltiplas Instâncias (MIL). Os autores também empregaram técnicas de *data augmentation* baseadas em rotações e *flips*. Os resultado foram 90% de acurácia em 20x e 82% em 40x.

O trabalho de [Vailankanni and Nagarajan 2026] propõem uma abordagem baseada em aprendizado profundo para classificação. A metodologia combina uma CNN para extração de características com técnicas de aprendizado autossupervisionado, usando o

dataset LungHist700. Em seguida, aplica seleção de atributos via MRMR e utiliza uma *Graph CNN* como classificador. Os resultados mostram 97% de acurácia.

[Sarita et al. 2025] apresentam um método de fusão de características com arquitetura dupla para a classificação. A abordagem combina uma *ConvNeXt* com *Data-efficient Image Transformer* (DeiT). Os experimentos foram conduzidos nos *datasets LC25000* [Borkowski et al. 2019] e *LungHist700*, alcançando acurácia de 99,71% no subconjunto de pulmão do *LC25000*, 99,95% no subconjunto de cólon e 89,45% no *LungHist700*.

Embora os estudos apresentados tenham resultados promissores, observa-se que a maioria das abordagens está concentrada na proposição de arquiteturas ou estratégias específicas. Havendo uma carência de análises comparativas envolvendo diferentes modelos de DL, da mesma forma, o impacto de técnicas de aumento de dados e de estratégias de *ensemble learning* ainda permanece pouco explorado nesse conjunto de dados. Diante desse cenário, este trabalho investiga o desempenho de diferentes modelos de DL, otimização de hiperparâmetros e aumento de dados. Adicionalmente, são avaliadas técnicas de *ensemble learning*, buscando aprimorar o desempenho na classificação de imagens histopatológicas, uma vez que, os métodos estudados podem ser complementares.

3. Metodologia

Para alcançar o objetivo deste estudo, a metodologia foi organizada conforme ilustrado na Figura 1.

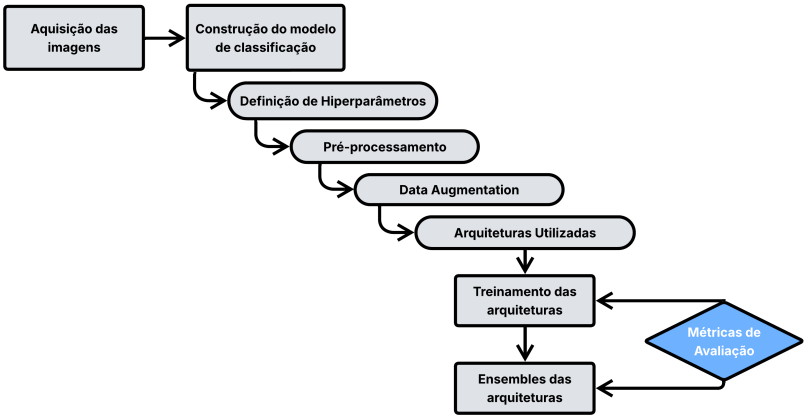
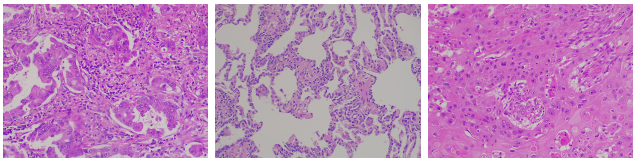


Figura 1. Fluxograma das etapas da metodologia proposta.

3.1. Aquisição das imagens

Neste trabalho, foi utilizado o *dataset LungHist700* [Diosdado et al. 2024], contendo 691 imagens histopatológicas de pulmão, provenientes de 45 pacientes. As imagens estão distribuídas em três tipos: adenocarcinoma com 280 imagens, carcinoma de células escamosas tendo 260 imagens e tecido normal com 151 imagens. Todas as imagens foram obtidas a partir de lâminas histológicas fixadas em formalina e coradas com hematoxilina e eosina (H&E), nas ampliações de 20× e 40×. A Figura 2 apresenta exemplos das imagens utilizadas neste estudo.

Seguindo a estratégia proposta pelos autores do *dataset*, os dados foram divididos em 80% para treinamento, 10% para validação e 10% para teste. A partição foi realizada no nível de paciente e mantida para ambas as ampliações, garantindo que imagens de um mesmo paciente não fossem compartilhadas entre os conjuntos. A adoção desse protocolo experimental visa preservar a consistência metodológica com a literatura, favorecendo a reprodutibilidade dos resultados e a comparação objetiva entre diferentes abordagens.



(I) (II) (III)

Figura 2. Imagens do dataset LungHist700 com ampliação de 20x: (I) adenocarcinoma, (II) tecido normal e (III) carcinoma de células escamosas.

3.2. Construção do modelo de classificação

Nesta seção, apresentamos os passos desenvolvidos para construção do modelo proposto neste trabalho.

3.2.1. Definição de Hiperparâmetros de treinamento

Inicialmente, foi conduzida uma etapa de otimização de hiperparâmetros por meio de *grid search*, na qual foram avaliadas 48 combinações de parâmetros. Para garantir a reprodutibilidade, foi utilizada a semente aleatória 42 em todas as etapas que envolvidas. Essa estratégia é empregada em tarefas de *machine learning* (ML) e *deep learning* para identificar configurações capazes de melhorar o desempenho dos modelos [Jiang and Xu 2022]. Para essa análise preliminar, foi adotada a arquitetura *VGG-16* [Simonyan and Zisserman 2014], inicializada com pesos pré-treinados no banco de dados *ImageNet*[1]. A escolha desse modelo fundamenta-se em sua ampla utilização em estudos de classificação de imagens médicas e em seu menor custo computacional [Klangbunrueang et al. 2025]. O objetivo dessa etapa foi definir uma configuração inicial de treinamento, posteriormente empregada nos experimentos conduzidos com as demais arquiteturas avaliadas. A Tabela 1 apresenta a combinação de hiperparâmetros selecionada. Com base nessa configuração, os modelos foram treinados por 100 épocas, com *early stopping* de 15 épocas para todas as arquiteturas avaliadas.

Tabela 1. Configurações de treinamento e espaço de busca.

Parâmetro	Selecionado	Busca
Imagem	224×224	60, 160, 224
Batch Size	32	32
Otimizador	AdamW	AdamW
Learning Rate	3×10^{-4}	1×10^{-4}, 3×10^{-4}
Weight Decay	1×10^{-4}	1×10^{-5}, 1×10^{-4}
Loss	CrossEntropyLoss	CrossEntropyLoss
Scheduler	CosineAnnealingLR	None, CosineAnnealingLR
Dropout	0.3	0.0, 0.3

[1]https://www.image-net.org/

3.2.2. Pré-Processamento

Após a delimitação dos hiperparâmetros apresentados na Seção 3.2.1, as imagens foram redimensionadas para 224 $\times$ 224, de modo a padronizar as dimensões de entrada dos modelos. Em seguida, foram normalizadas utilizando os parâmetros do conjunto *ImageNet* (média = [0,485; 0,456; 0,406] e desvio padrão = [0,229; 0,224; 0,225]), com o objetivo de uniformizar os valores dos *pixels* e adequá-los à distribuição esperada pelos modelos.

3.2.3. Data Augmentation

Para mitigar o desbalanceamento do *dataset*, foi utilizada a técnica de *data augmentation*, com o objetivo de aumentar a variabilidade dos dados e melhorar a capacidade de generalização do modelo. Para isso, foi aplicado o método *RandAugment* [Cubuk et al. 2020], uma estratégia de aumento de dados baseada na aplicação aleatória de transformações em imagens.

O método considera um conjunto de K transformações possíveis, como rotação, *flip*, ajuste de contraste e cor, sendo cada operação selecionada com probabilidade uniforme ($\frac{1}{K}$). Diferentemente de abordagens tradicionais de aumento de dados, o *RandAugment* simplifica o processo ao utilizar apenas dois hiperparâmetros: a magnitude das transformações (m) e o número de operações aplicadas sequencialmente (n) [Faryna et al. 2021]. O *RandAugment* tem apresentado bons resultados envolvendo imagens histopatológicas com H&E [Faryna et al. 2024], contribuindo para a melhoria da robustez e da capacidade de generalização. Neste trabalho, foram utilizados $n = 2$ e $m = 5$, permitindo a aplicação de duas transformações aleatórias com intensidade moderada em cada imagem do conjunto de treinamento. Para complementar a estratégia, foi utilizado o *WeightedRandomSampler* da biblioteca *PyTorch*[2]. Nesse método, cada amostra do conjunto de treinamento recebe um peso associado à sua classe. Esses pesos determinam a probabilidade de uma amostra ser selecionada durante o processo de treinamento: quanto maior o peso, maior a chance de a amostra ser escolhida. Dessa forma, amostras pertencentes a classes minoritárias são selecionadas com maior frequência, promovendo uma distribuição mais equilibrada dos dados apresentados ao modelo.

3.2.4. Arquiteturas Utilizadas no Treinamento

A partir da definição dos hiperparâmetros apresentados na Tabela 1, foram avaliadas diferentes arquiteturas de DL amplamente empregadas em tarefas de classificação de imagens médicas. A seleção dos modelos buscou contemplar distintos paradigmas arquiteturais, permitindo uma análise comparativa entre arquiteturas convolucionais clássicas, modelos modernos e abordagens baseadas em mecanismos de atenção.

Entre os modelos utilizados estão a *ResNet-50* [He et al. 2016], baseada em conexões residuais; a *DenseNet-121* [Huang et al. 2017], caracterizada pelo compartilhamento intensivo de informações entre camadas; a *EfficientNet-B3* [Tan and Le 2019], projetada para otimizar a relação entre desempenho e custo computacional; a *ConvNeXt-*

[2]https://pytorch.org/

Tiny [Liu et al. 2022], modelo inspirado nos ViTs; e a *Vision Transformer-B16 (ViT-B16)* [Dosovitskiy et al. 2020], baseada em mecanismos de autoatenção. Todos os modelos foram inicializados com pesos pré-treinados no *ImageNet*, usando *transfer learning*. Além disso, para assegurar uma comparação justa entre as modelos, todos os experimentos foram conduzidos utilizando os mesmos hiperparâmetros apresentados na Tabela 1.

3.3. Ensemble Learning

Nos últimos anos, diversos estudos têm demonstrado que o uso de *ensembles* proporciona um aprimoramento significativo na taxa de acerto em sistemas de classificação, quando comparados com modelos individuais [Mienye and Sun 2022]. A ideia fundamental por trás dessa técnica baseia-se no fato de que modelos isolados possuem limitações e estão sujeitos a erros. A seguir, são apresentadas as abordagens de *ensemble* utilizadas.

3.3.1. Hard Voting

Essa técnica integra as predições de múltiplos modelos por meio de votação majoritária. Cada modelo fornece apenas a classe prevista, e aquela que recebe o maior número de votos é selecionada como resultado final [Krishna and Reka 2025], conforme a Figura 3.

3.3.2. Soft Voting

Essa abordagem utiliza as probabilidades previstas pelos modelos, calculando a média entre elas. A classe que apresenta a maior probabilidade média é escolhida como predição final [Krishna and Reka 2025], de acordo com a Figura 3.

3.3.3. Weighted Voting

Essa estratégia agrega as predições de múltiplos modelos, atribuindo pesos proporcionais ao desempenho de cada um. Nesse caso, o *F1-Score* obtido no treinamento é utilizado como peso de cada modelo. As probabilidades previstas são multiplicadas pelos respectivos pesos e, posteriormente, somadas, sendo selecionada como resultado final a classe que apresenta a maior probabilidade ponderada [Krishna and Reka 2025].

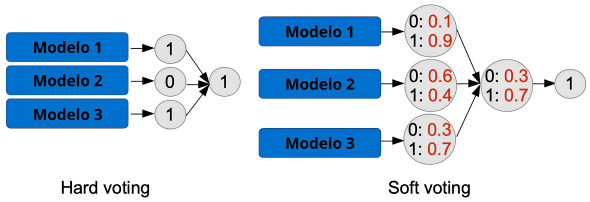


Figura 3. Estratégias de *ensemble* utilizadas.

3.4. Métricas de Avaliação

Para a avaliação dos resultados, foram utilizadas métricas, como: Acurácia, *Kappa*, Precisão, *Recall*, *F1-score* e Matriz de Confusão [Sammut and Webb 2011].

4. Resultados e Discussões

Nesta seção, são apresentados os resultados obtidos na avaliação de desempenho dos modelos e das técnicas empregadas. Para analisar a capacidade das arquiteturas e compreender o comportamento dos modelos, foram considerados três cenários: (i) sem *data augmentation*, (ii) com *data augmentation* e (iii) por meio de estratégias de *ensemble*.

A Tabela 2 apresenta os resultados obtidos sem a aplicação das técnicas de *data augmentation* nas ampliações de 20x, onde o ViT-B16 apresentou desempenho superior aos demais modelos. Esse resultado sugere que o mecanismo de autoatenção foi capaz de explorar eficientemente informações globais presentes nas imagens de menor ampliação, nas quais uma área maior do tecido é observada. Em contraste, o *ConvNeXt-Tiny* apresentou os menores resultados, indicando dificuldades para aprender representações discriminativas sem o auxílio do aumento de dados.

Tabela 2. Resultados dos modelos no conjunto 20x sem RandAugment.

Modelo	Acc	Kap.	Prec.	Rec.	F1
ResNet-50	62.50	43.86	72.84	66.76	62.38
EfficientNet-B3	47.92	24.58	43.65	54.28	42.74
DenseNet-121	47.92	19.30	70.82	48.29	43.58
ViT-B16	**83.33**	**74.59**	**85.96**	**85.87**	**84.53**
ConvNeXt-Tiny	37.50	00.00	12.50	33.33	18.18

Para o conjunto 40x, os resultados apresentados na Tabela 3 a *DenseNet-121* destacou-se como o modelo mais eficaz. Isso pode está ligado a suas conexões densas, que favoreceram o reaproveitamento de características relevantes para a identificação de detalhes celulares e nucleares, mais evidentes nessa ampliação. Os resultados também mostram que o desempenho das arquiteturas varia conforme a escala de análise, sugerindo que diferentes modelos podem se beneficiar de características específicas presentes em cada nível de ampliação das imagens histopatológicas.

Tabela 3. Resultados dos modelos no conjunto 40x sem RandAugment.

Modelo	Acc	Kap.	Prec.	Rec.	F1
ResNet-50	56.00	31.21	41.59	50.78	43.89
EfficientNet-B3	64.00	43.07	43.37	57.45	49.12
DenseNet-121	**66.00**	**48.70**	**67.79**	**67.53**	**64.26**
ViT-B16	56.00	31.97	47.86	51.67	43.89
ConvNeXt-Tiny	40.00	00.00	13.33	33.33	19.05

Já as Tabelas 4 e 5 exibem os resultados obtidos com a aplicação das tecnicas de *data augmentation* nas ampliações de 20x e 40x. Em comparação aos experimentos sem aumento de dados, observa-se uma melhora expressiva em todas as métricas avaliadas. No conjunto 20x, por exemplo, o *DenseNet-121* elevou sua acurácia de 47,92% para 95,83%, enquanto o *ConvNeXt-Tiny* passou de 37,50% para 83,33%. Resultados semelhantes foram observados no conjunto 40x, evidenciando o impacto positivo do aumento de dados na capacidade de generalização dos modelos.

Entre as arquiteturas avaliadas, o *DenseNet-121* apresentou o desempenho mais consistente, alcançando os melhores resultados em ambos ampliações. Esse comportamento pode estar relacionado às suas conexões densas, conforme discutido anteriormente. Além disso, o ganho observado após a aplicação do *RandAugment* e do *Weigh-*

tedRandomSampler sugere que a maior diversidade de amostras geradas e o balanceamento proporcionado pelos pesos de amostragem permitiram uma aprendizagem mais robusta dos padrões histopatológicos, reduzindo o sobreajuste e melhorando a capacidade de generalização dos modelos.

Tabela 4. Resultados dos modelos no conjunto 20x com RandAugment.

Modelo	Acc	Kap.	Prec.	Rec.	F1
ResNet50	**95.83**	93.57	**96.67**	95.22	95.76
EfficientNet-B3	91.67	87.26	91.75	92.88	92.14
DenseNet121	**95.83**	**93.63**	95.47	**96.49**	**95.80**
ViT-B16	85.42	77.40	88.31	85.96	86.58
ConvNeXt-Tiny	83.33	74.77	83.37	85.58	83.83

Tabela 5. Resultados dos modelos no conjunto 40x com RandAugment.

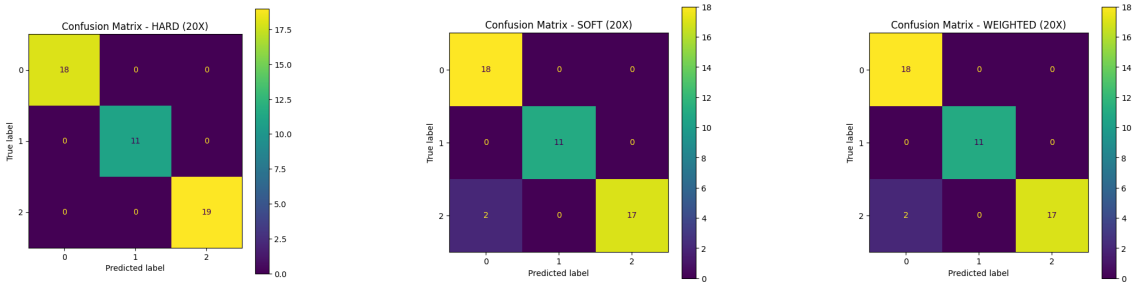
Modelo	Acc	Kap.	Prec.	Rec.	F1
ResNet-50	94.00	90.81	94.89	93.51	94.05
EfficientNet-B3	94.00	90.88	93.73	93.81	93.72
DenseNet-121	**96.00**	**93.92**	**95.58**	**95.77**	**95.63**
ViT-B16	88.00	81.79	89.29	87.91	88.03
ConvNeXt-Tiny	92.00	87.92	91.37	92.14	91.52

Para compor o *ensemble*, foram selecionados os três modelos com melhor desempenho individual após a aplicação do *RandAugment*: *DenseNet-121*, *ResNet-50* e *EfficientNet-B3*. A seleção considerou os maiores valores de Acurácia e *F1-Score*, além da diversidade arquitetural entre os modelos. Optou-se por utilizar apenas os modelos de melhor desempenho para evitar que preditores menos precisos introduzissem erros adicionais no processo de votação. Dessa forma, a composição do *ensemble* prioriza a combinação de modelos de alto desempenho e arquiteturas distintas, em vez da inclusão indiscriminada de todos os modelos avaliados.

Os resultados das estratégias de *ensemble* são apresentados nas Tabelas 6 e 7, com as respectivas matrizes de confusão nas Figuras 4 e 5. No conjunto 20x, o *Hard Voting* destacou-se ao classificar corretamente todas as amostras. Esse resultado deve ser interpretado considerando o número reduzido de pacientes do *dataset* e a divisão dos dados no nível de paciente, sem o risco de *data leakage*. Em contraste, o *Soft* e o *Weighted Voting* obtiveram 95,83%, errando duas amostras da classe normal. Essa diferença pode ser atribuída ao comportamento dos modelos base: quando um modelo erra com alta confiança, sua probabilidade elevada distorce a média e suprime o voto correto da maioria, algo que o *Hard Voting* evita ao tratar todos os votos igualmente.

Tabela 6. Resultados das estratégias de *ensemble* no conjunto 20x.

Ensemble	Acc	Kap.	Prec.	Rec.	F1
Hard Voting	**100**	**100**	**100**	**100**	**100**
Soft Voting	95.83	93.60	96.67	96.49	96.39
Weighted Voting	95.83	93.60	96.67	96.49	96.39



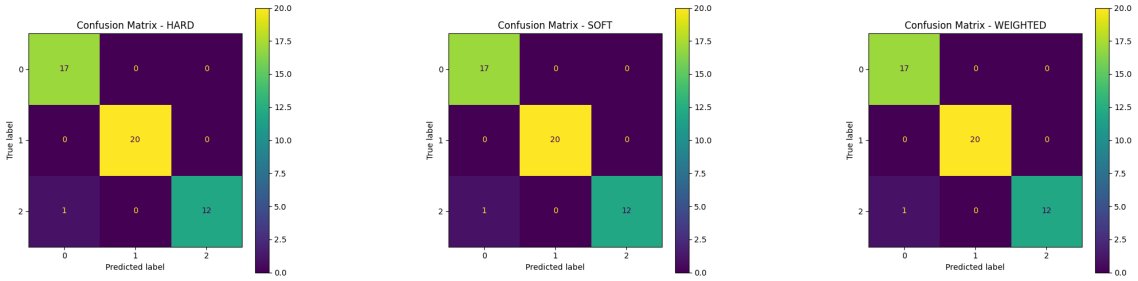
(I) Hard Voting (II) Soft Voting (III) Weighted Voting

Figura 4. Matrizes de confusão dos métodos de ensemble para o conjunto 20x.

Para o conjunto 40x, a Tabela 7 e a Figura 5 mostram que os três métodos convergiram para o mesmo desempenho, alcançando 98% de acurácia e cometendo o mesmo erro: a amostra de índice 44, pertencente à classe normal, foi classificada como adenocarcinoma. Isso ocorreu porque os modelos individuais *ResNet-50* e *EfficientNet-B3* cometeram esse erro, enquanto apenas a *DenseNet-121* realizou a classificação correta. Como dois dos três modelos concordaram na predição incorreta, nenhuma estratégia de agregação foi capaz de reverter o erro, evidenciando que a eficácia do *ensemble* depende fundamentalmente da diversidade entre os modelos utilizados.

Tabela 7. Resultados das estratégias de *ensemble* no conjunto 40x.

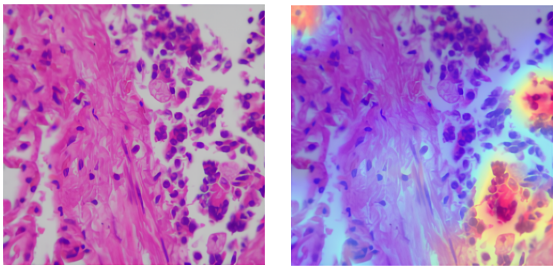
Ensemble	Acc	Kap.	Prec.	Rec.	F1
Hard Voting	98.00	96.95	98.15	97.44	97.71
Soft Voting	98.00	96.95	98.15	97.44	97.71
Weighted Voting	98.00	96.95	98.15	97.44	97.71



(I) Hard Voting (II) Soft Voting (III) Weighted Voting

Figura 5. Matrizes de confusão dos ensembles para o conjunto 40x.

A Figura 6 apresenta a amostra de índice 44 e o mapa *Grad-CAM*, que mostra que as regiões que mais influenciaram a decisão do modelo estão concentradas próximas às bordas da imagem. Como essas áreas ocupam uma porção reduzida da amostra e encontram-se afastadas da região central, os modelos tiveram maior dificuldade em interpretar corretamente o contexto global do tecido, atribuindo maior relevância a padrões celulares localizados que apresentam semelhanças visuais com características observadas em adenocarcinomas.



(I) Imagem original (II) Grad-CAM

Figura 6. Comparação entre a imagem original e Grad-CAM correspondente.

Na Tabela 8, apresentamos uma comparação com os demais métodos listados anteriormente. Os resultados obtidos em nosso método foram competitivos em relação aos demais. Os modelos individuais alcançaram acurácias de até 95% em 20x e 96% em 40x, superando os resultados reportados por [Diosdado et al. 2024] e [Sarita et al. 2025]. Além disso, o método *Hard Voting* apresentou o melhor desempenho geral, atingindo 100% de acurácia no conjunto 20x e 98,0% no conjunto 40x, evidenciando a eficácia das estratégias de *ensemble* para a classificação.

Tabela 8. Comparação entre resultados da literatura usando *LungHist700*.

Método	**Ampliação**	**Acc**
ResNet50 [Diosdado et al. 2024]	20x	90.00
ResNet50 [Diosdado et al. 2024]	40x	82.00
MuSSL + mRMR + VGCN [Vailankanni and Nagarajan 2026]	–	97.00
ConvNeXt + DeiT [Sarita et al. 2025]	–	89.45
DenseNet121 + RandAugment	20x	95.83
DenseNet121 + RandAugment	40x	96.00
Ensemble (Hard Voting)	20x	**100**
Ensemble (Hard Voting)	40x	**98.00**

5. Conclusão

Neste trabalho, foram investigadas diferentes arquiteturas de DL para a classificação de imagens histopatológicas de câncer de pulmão, combinando estratégias de aumento de dados e *ensemble learning*. Os experimentos demonstraram que o *RandAugment* aliado ao *WeightedRandomSampler* contribuiu para o desempenho dos modelos, com ganhos expressivos em relação ao cenário sem aumento de dados. Entre as arquiteturas avaliadas, a *DenseNet-121* apresentou o comportamento mais consistente nas ampliações, enquanto as estratégias de *ensemble* superaram os resultados individuais. Embora o *Hard Voting* tenha alcançado 100% de acurácia no conjunto 20x, esse resultado deve ser interpretado com cautela devido ao pequeno tamanho do conjunto de teste. Para trabalhos futuros, pretende-se: i) realizar validação externa utilizando a base *LC25000*, fim de avaliar a capacidade de generalização do estudo em diferentes cenários; ii) explorar outras técnicas de *data augmentation*, visando aprimorar a generalização.

6. Agradecimentos

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Código de Financiamento 001, da Fundação de

Amparo à Pesquisa do Estado do Piauí (FAPEPI) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Referências

Acs (2024). Cancer a-z. about lung cancer. https://www.cancer.org/cancer/types/lung-cancer/about/what-is.html.

Borkowski, A. A., Bui, M. M., Thomas, L. B., Wilson, C. P., DeLand, L. A., and Mastorides, S. M. (2019). Lung and colon cancer histopathological image dataset (lc25000). *arXiv preprint arXiv:1912.12142*.

Cubuk, E. D., Zoph, B., Shlens, J., and Le, Q. V. (2020). Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703.

Diosdado, J., Gilabert, P., Seguí, S., and Borrego, H. (2024). Lunghist700: A dataset of histological images for deep learning in pulmonary pathology. *Scientific Data*, 11(1):1088.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Faryna, K., van der Laak, J., and Litjens, G. (2021). Tailoring automated data augmentation to hamp;e-stained histopathology. In Heinrich, M., Dou, Q., de Bruijne, M., Lellmann, J., Schläfer, A., and Ernst, F., editors, *Proceedings of the Fourth Conference on Medical Imaging with Deep Learning*, volume 143 of *Proceedings of Machine Learning Research*, pages 168–178. PMLR.

Faryna, K., van der Laak, J., and Litjens, G. (2024). Automatic data augmentation to improve generalization of deep learning in he stained histopathology. *Computers in Biology and Medicine*, 170:108018.

Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., and Yener, B. (2009). Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2:147–171.

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.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.

Iarc (2022). Population factsheets - worldwide. https://gco.iarc.fr/today/en/fact-sheets-populations#global.

Inca (2022). *Estimativa 2023: incidência de câncer no Brasil*. INCA, Rio de Janeiro.

Jiang, X. and Xu, C. (2022). Deep learning and machine learning with grid search to predict later occurrence of breast cancer metastasis using clinical data. *Journal of clinical medicine*, 11(19):5772.

Klangbunrueang, R., Pookduang, P., Chansanam, W., and Lunrasri, T. (2025). Ai-powered lung cancer detection: Assessing vgg16 and cnn architectures for ct scan image classification. *Informatics*, 12(1):18.

Krishna, S. S. and Reka, S. S. (2025). Early detection of retinopathy of prematurity using voting classifier-based ensemble deep learning models. *International Journal of Computational Intelligence Systems*, 18(1):238.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986.

Mienye, I. D. and Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, 10:99129–99149.

Myers, D. J. and Wallen, J. M. (2023). Lung adenocarcinoma. In *StatPearls [Internet]*. StatPearls Publishing.

Relli, V., Trerotola, M., Guerra, E., and Alberti, S. (2019). Abandoning the notion of non-small cell lung cancer. *Trends in molecular medicine*, 25(7):585–594.

Sammut, C. and Webb, G. I. (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.

Sarita, Shukla, P. K., Dhaka, V. S., and Jindal, N. (2025). Dual-backbone feature extraction framework for lung cancer classification in histopathology images. *International Journal of Information Technology*, pages 1–18.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Takahashi, S., Sakaguchi, Y., Kouno, N., Takasawa, K., Ishizu, K., Akagi, Y., Aoyama, R., Teraya, N., Bolatkan, A., Shinkai, N., et al. (2024). Comparison of vision transformers and convolutional neural networks in medical image analysis: a systematic review. *Journal of Medical Systems*, 48(1):84.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.

Vailankanni, N. and Nagarajan, B. (2026). Deep feature optimization using fusion of multiple self-supervised learning approaches and filter-based feature selection for lung cancer histopathology classification. *PloS one*, 21(4):e0348194.