

Avaliando o Impacto de Técnicas de Aumentos de Dados em Modelos com Arquiteturas Diferentes

Victor Hugo Lara Souza

Sistemas de Informação

Universidade Federal de Viçosa - Campus Rio Paranaíba

Rio Paranaíba, Brasil

victor.h.souza@ufv.br

Abstract—Brain tumor classification from magnetic resonance imaging (MRI) plays a crucial role in supporting early diagnosis and improving treatment. Recent advances in deep learning have enabled the development of highly accurate automated systems capable of assisting medical specialists in the interpretation of complex visual patterns. In this work, we evaluate the performance of two modern architectures, ResNet-18, a CNN model, and ViT-B/16, a Vision Transformer for classifying images on the Brain Tumor MRI dataset, consisting of 7023 images. Investigating the impact of data augmentation on both architectures. The experiments show that ResNet-18 benefits from augmentation, achieving improved accuracy and generalization, while ViT-B/16 exhibits a slight performance decrease when augmented data is applied. Overall, ViT-B/16 attained the highest performance without augmentation, reaching 98.51% accuracy, while ResNet-18 achieved its best result of 97.79% with augmented training. These results highlight important architectural differences and provide insights regarding the suitability of CNNs and Vision Transformers for medical image classification tasks.

Index Terms—Visão Computacional, Redes Neurais Convolucionais, Imagens Médicas, aumento de dados

I. INTRODUÇÃO

A cada ano, centenas de milhares de pessoas no mundo recebem o diagnóstico de tumores, uma condição que, embora relativamente rara, gera impacto clínico e social. De acordo com dados do GLOBOCAN 2022, foram estimados aproximadamente 321.600 novos casos de câncer no cérebro e sistema nervoso central, com cerca de 248.400 mortes atribuídas à doença no mesmo ano [GloboCan, 2023,]. No Brasil, o Instituto Nacional de Câncer estima cerca de 11.490 novos casos anuais de tumores cerebrais para o triênio 2023–2025 [Instituto Oncoguia, 2025,]. Além disso, a carga global desses tumores vem crescendo: entre 1990 e 2021, a prevalência ajustada por idade aumentou mais de 120% [PMC, 2025,]. Esses números ressaltam a urgência de métodos mais eficazes para diagnóstico precoce, classificação e tratamento, especialmente em regiões com recursos limitados.

Enquanto isso, nos últimos anos, os avanços em visão computacional e deep learning têm desempenhado um papel fundamental na área médica, especialmente no suporte ao diagnóstico por imagem. Entre as aplicações de maior impacto, destaca-se a classificação de tumores cerebrais, uma tarefa crítica para o tratamento precoce e para o aumento da taxa de sobrevivência dos pacientes. Tradicionalmente, esse processo depende da análise visual realizada por médicos especialistas,

o que pode ser demorado, sujeito a variabilidade entre observadores e limitado pelo volume crescente de exames clínicos.

Nesse contexto, o presente trabalho investiga a aplicação de técnicas de visão computacional para a classificação de tumores cerebrais em imagens de ressonância magnética, explorando modelos de aprendizado profundo capazes de extrair características mesmo em condições visuais adversas e avaliando o impacto de estratégias de aumento de dados em modelos com arquiteturas diferentes. O objetivo é avaliar o desempenho de abordagens modernas e discutir suas contribuições, limitações e potencial para uso clínico.

II. REVISÃO BIBLIOGRÁFICA

A classificação automatizada de tumores cerebrais em imagens de ressonância magnética (MRI) tem recebido crescente atenção na literatura, impulsionada pelo avanço dos métodos de aprendizagem profunda e pela necessidade de diagnósticos mais rápidos e precisos. Entre as abordagens mais empregadas, destacam-se as Redes Neurais Convolucionais (CNNs), que têm se consolidado como a técnica predominante em visão computacional e os Vision Transformers.

Nesse sentido, [SILVA et al., 2024] realizou um estudo comparativo entre uma CNN tradicional e uma CNN enriquecida com um mecanismo de self-attention. Nesse mecanismo, as matrizes query, key e value são utilizadas para gerar pesos atencionais que destacam regiões mais relevantes para a classificação dos tumores, aumentando a capacidade da rede de focar seletivamente em estruturas importantes. Utilizando o dataset "Brain Tumor MRI", com 7023 imagens distribuídas em quatro classes (glioma, meningioma, pituitário e sem tumor), o autor demonstrou que a adição do módulo de atenção elevou o desempenho global do modelo, com ganhos de 1,98% em acurácia, 2,07% em precisão, 2,18% em sensibilidade e 1,72% em F1-score em relação à CNN base. O impacto foi ainda mais significativo na classe meningioma, cuja acurácia aumentou em 6,54%, evidenciando que o mecanismo de atenção contribuiu para melhorar a sensibilidade a padrões difíceis de distinguir no modelo convencional.

Já [Alemayehu, 2025] propõe uma abordagem diferente, o estudo pretende desenvolver uma arquitetura CNN leve para classificação de tumores cerebrais em imagens de ressonância magnética (MRI), utilizando o dataset "Brain Tumor MRI" para classificar as classes: glioma, meningioma, pituitary e

notumor. A CNN desenvolvida do zero alia baixo custo computacional a uma acurácia competitiva de 98,78%. Essa proposta destaca que, com pré-processamento adequado, data augmentation e otimização sistemática de hiperparâmetros, é possível obter desempenho próximo ao de modelos pré-treinados.

[Gómez-Guzmán et al., 2023] utiliza o mesmo dataset com uma abordagem híbrida, ele propõe utilizar a arquitetura MobileNetV2 para a extração de características enquanto utiliza classificadores tradicionais, nesse caso o Support Vector Machine (SVM) para a classificação. Os resultados mostraram que o modelo não apenas apresenta boas métricas como reduz o trabalho computacional.

Uma comparação entre diferentes arquiteturas é feita por [Disci et al., 2025], os autores compararam as redes Xception, MobileNetV2, InceptionV3, ResNet50, VGG16, e DenseNet121, que passaram por fine-tuning com transfer-learning com pré-processamento e aumento de dados. A arquitetura Xception se destacou com uma acurácia de 98.73% e F1 score de 95.29%.

III. MATERIAIS E MÉTODOS

A. Dataset

O dataset utilizado foi o "Brain Tumor MRI", um dataset para a classificação de imagens de tumor cerebral em ressonância magnética. O dataset está disponível no Kaggle, ele possui 7023 imagens divididas em 4 classes: glioma, meningioma, no tumor e pituitary. Este dataset se trata da junção de outros 3 datasets: figshare, SARTAJ dataset e Br35H. O dataset vem originalmente dividido apenas em treino e teste, com 5712 e 1311 respectivamente. Para este estudo foram utilizadas todas as imagens de todas as classes. A figura 1 apresenta amostras do dataset:

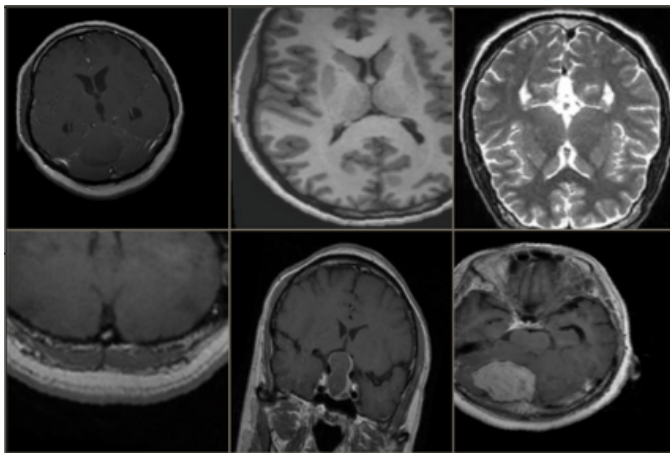


Fig. 1: Exemplo de imagens contidas no dataset.

B. Pré-Processamento

O pré-processamento das imagens teve como objetivo padronizar a entrada do modelo e garantir compatibilidade com arquiteturas pré-treinadas no ImageNet, além de preparar

os dados para o processo de data augmentation. Todas as etapas foram implementadas utilizando o Pytorch.

Inicialmente, cada imagem foi carregada e convertida explicitamente para o formato RGB. Em seguida, as imagens foram submetidas a operações de redimensionamento por corte (crop) para alcançar o tamanho de entrada esperado pelos modelos utilizados.

Para o conjunto de treinamento, adotou-se duas estratégias, uma com aumento de dados e outra sem, na versão com aumento de dados foi aplicado RandomResizedCrop com resolução final de 224×224 pixels, na versão sem o aumento de dados foi aplicado o CenterCrop com a mesma resolução. Já para os conjuntos de validação e teste, aplicou-se apenas o CenterCrop(224). Como etapa final, todas as imagens foram normalizadas utilizando a média e o desvio padrão do conjunto ImageNet, procedimento padrão ao se utilizar modelos pré-treinados.

C. Aumento de Dados

Para reduzir o risco de overfitting e aumentar a capacidade de generalização do modelo, foram adotadas duas estratégias de transformações nas imagens, na primeira com aumento de dados aplicou-se operações de transform por meio do módulo torchvision.transforms do Pytorch.

Foram aplicados as seguintes transformações de forma probabilística: RandomHorizontalFlip(espelhamento horizontal aplicado com probabilidade de 50%), RandomRotation (rotações aleatórias dentro do intervalo de $\pm 15^\circ$), ColorJitter (alterações aleatórias de brilho, contraste, saturação e matiz), GaussianBlur (aplicação de desfoque gaussiano com kernel-size=3). Todas essas operações foram aplicadas antes da conversão da imagem para tensor e da normalização correspondente ao conjunto ImageNet.

Na versão dos experimentos sem aumento de dados foram aplicadas apenas o CenterCrop, como descrito na subseção de pré-processamento.

D. Divisão dos Dados

A divisão do conjunto de dados foi realizada com o objetivo de garantir uma avaliação confiável do modelo e prevenir vazamento de informação entre as etapas de treinamento e teste. O processo foi conduzido utilizando a função train-test-split da biblioteca Scikit-learn.

O conjunto de imagens original vem com duas pastas, treino e teste, cada uma com 4 pastas para cada classe. Para o experimento a organização foi modificada para possuir apenas 4 pastas no segundo nível, correspondentes às classes glioma, meningioma, pituitary e notumor, com todo o conteúdo de imagens correspondente dentro delas.

O conjunto foi dividido na razão 60/20/20 para treino, validação e teste respectivamente, essa divisão foi realizada em duas etapas sucessivas. A primeira separou 20% dos dados para o conjunto de teste. A segunda etapa dividiu os 80% restantes em treinamento e validação, garantindo que o conjunto de validação também representasse 20% do total original. O processo utilizou a semente de reprodutibilidade SEED =

42, garantindo que os subconjuntos fossem consistentes entre execuções.

E. Arquiteturas

Nesta pesquisa foram empregados dois modelos para tarefas de classificação de imagens: a **ResNet-18**, baseada em convoluções profundas com conexões residuais, e o **ViT-B/16**, representando a classe de modelos que utilizam mecanismos de atenção originalmente desenvolvidos para processamento de linguagem natural. A seguir, estão as principais características de cada arquitetura:

A ResNet-18 [He et al., 2016] é uma rede neural convolucional pertencente à família das *Residual Networks*. Seu principal diferencial é o uso de *skip connections*, que permitem a propagação direta do gradiente entre camadas, facilitando o treinamento de redes profundas e reduzindo o problema do *vanishing gradient*. A arquitetura utilizada neste trabalho foi inicializada com pesos pré-treinados no conjunto ImageNet, permitindo a aplicação da técnica de *fine-tuning*.

O ViT-B/16 [Dosovitskiy, 2020] utiliza o paradigma dos *Transformers* para visão computacional, dispensando o uso de convoluções. A imagem de entrada é dividida em patches de tamanho 16×16 , que são linearmente projetados em embeddings vetoriais. Esses embeddings, juntamente com um token de classificação, são então processados por uma pilha de camadas Transformer compostas por mecanismos de autoatenção multi-cabeças e camadas feed-forward posicionais. Assim como a ResNet-18, o ViT-B/16 também foi inicializado com pesos pré-treinados, permitindo a adaptação das representações aprendidas por meio de *fine-tuning*.

F. Hiperparâmetros e Configurações

Para o treinamento, foi adotado um batch size igual a 64. A taxa de aprendizado inicial foi definida como 0.0001, adequada ao fine-tuning de modelos pré-treinados, reduzindo o risco de alterações abruptas nos pesos da rede. O otimizador selecionado foi o AdamW, escolhido por apresentar melhor regularização e comportamento mais estável em arquiteturas profundas quando comparado ao Adam tradicional.

Para controle dinâmico da taxa de aprendizado durante o treinamento, utilizou-se o agendador CosineAnnealingLR, configurado para atuar ao longo de todas as 50 épocas previstas. Esse scheduler ajusta a taxa de aprendizado seguindo uma curva cosseno, reduzindo-a gradualmente até valores próximos de zero, o que favorece a estabilização da solução nas últimas épocas do treinamento.

A função de perda empregada foi a CrossEntropyLoss, adequada a problemas de classificação multiclasse como o "Brain Tumor MRI". Adicionalmente, todas as sementes aleatórias foram fixadas em SEED = 42 para garantir reprodutibilidade dos resultados.

G. Treino

O processo de treinamento foi conduzido seguindo uma rotina estruturada em épocas (epochs) e ciclos de otimização que combinam propagação direta, cálculo de erro e backpropagation do gradiente.

Para cada lote, realiza-se forward pass, no qual as imagens são processadas pelo modelo e são produzidas as previsões correspondentes. Em seguida, calcula-se a função de perda por meio do critério CrossEntropyLoss. Após o cálculo do erro, procede-se ao backpropagation, atualizando os gradientes e posteriormente aplicando a etapa de otimização por meio do otimizador AdamW, que ajusta os pesos da rede de forma iterativa.

Durante o treinamento, registra-se o erro total e o número de previsões corretas, permitindo o cálculo da perda média e da acurácia ao final de cada época. Juntamente ao processo de otimização, o valor da taxa de aprendizado é ajustado dinamicamente pelo CosineAnnealingLR. Ao término de cada época, o modelo é avaliado no conjunto de validação.

H. Avaliação do Modelo

A avaliação do modelo foi realizada utilizando exclusivamente o conjunto de teste, separado previamente e não empregado em nenhuma etapa do treinamento ou validação.

Cada lote de imagens do conjunto de teste foi processado pelo modelo para gerar as respectivas previsões. A classe prevista foi obtida por meio da operação argmax sobre as saídas da rede. Além disso, foram calculadas as probabilidades associadas às previsões por meio da função softmax, permitindo análises mais detalhadas, como inspeção de confiança do modelo.

Com as previsões obtidas, foram geradas as métricas de desempenho utilizando as ferramentas da biblioteca Scikit-learn. A matriz de confusão foi construída para analisar os padrões de acertos e erros por classe, evidenciando possíveis dificuldades em categorias específicas. Em seguida, foi produzido o classification report, contendo métricas de acurácia, precision, recall e F1-score para cada classe, além das médias ponderadas.

I. Métricas

Aqui são apresentados os procedimentos utilizados para a avaliação dos modelos para a classificação de tumores cerebrais. O desempenho foi analisado a partir de métricas amplamente empregadas na literatura, incluindo acurácia, *precision*, *recall* e *F1-score*. Essas métricas permitem quantificar não apenas a proporção de acertos do modelo, mas também sua capacidade de identificar corretamente as instâncias positivas e negativas, aspecto fundamental em aplicações médicas.

A acurácia representa a proporção total de previsões corretas em relação ao conjunto de amostras avaliadas, sendo definida como:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1)$$

Embora forneça uma visão geral do desempenho, essa métrica pode ser insuficiente em cenários com desbalanceamento entre classes.

A métrica *precision* avalia a capacidade do modelo de evitar falsos positivos, indicando a fração de instâncias classificadas como positivas que realmente pertencem à classe positiva:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (2)$$

Em contextos clínicos, maior precisão reduz a probabilidade de alarmes incorretos.

O *recall* mede a proporção de amostras positivas que foram corretamente identificadas pelo modelo, sendo calculado por:

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (3)$$

Nesse tipo de aplicação, um alto *recall* é crucial, pois diminui a chance de casos positivos não serem detectados.

Por fim, o F1-score é definido como a média harmônica entre *precision* e *recall*, fornecendo uma métrica equilibrada entre ambos os aspectos:

$$\text{F1-score} = \frac{2TP}{2TP + FP + FN}. \quad (4)$$

Esse indicador é particularmente útil em situações de desbalanceamento entre classes.

J. Ambiente Computacional e Reprodutibilidade

Os experimentos foram realizados integralmente na plataforma Google Colab, que disponibiliza GPUs, para estes experimentos foi utilizado o ambiente NVIDIA Tesla T4. Para garantir reprodutibilidade, todas as sementes pseudoaleatórias foram definidas com o valor 42, abrangendo os módulos random, NumPy, PyTorch, ScikitLearn e Pandas.

IV. RESULTADOS E DISCUSSÃO

A. Geral

A seguir, são discutidos os resultados experimentais dos modelos avaliados, com destaque para o impacto do data augmentation. Os desempenhos foram analisados com base nas principais métricas utilizadas na literatura, permitindo uma comparação clara entre as arquiteturas ResNet-18 e ViT-B/16 em diferentes cenários de data augmentation arredondados para 4 dígitos após o ponto, os valores estão na tabela I:

TABLE I: Resultados dos Modelos, sendo resnet-noDA a arquitetura resnet18 na avaliação sem o aumento de dados e resnet18-DA a mesma arquitetura com aumento de dados, as mesmas siglas se aplicam ao vit nos dois casos.

Modelo	Acurácia	Precision	Recall	F1-Score
resnet18_noDA	0.9687	0.9684	0.9674	0.9678
resnet18_DA	0.9779	0.9776	0.9770	0.9772
vitb16_noDA	0.9851	0.9848	0.9846	0.9846
vitb16_DA	0.9829	0.9826	0.9825	0.9824

De modo geral, ambos os modelos obtiveram desempenhos elevados, com métricas superiores a 0.96 em todos os cenários. Entretanto, é possível analisar alguns detalhes. As figuras 2, 3, 4 e 5 permitem uma comparação visual e clara das diferenças entre as métricas obtidas:

A partir destes gráficos é possível perceber que o modelo vitb16 nos testes sem aumento de dados se destacou em todas as métricas, com uma acurácia de 0.9851, seguido pelo vitb16 com aumento de dados, resultado compatível com a literatura, estudos como [Asiri et al., 2023] mostram que arquiteturas

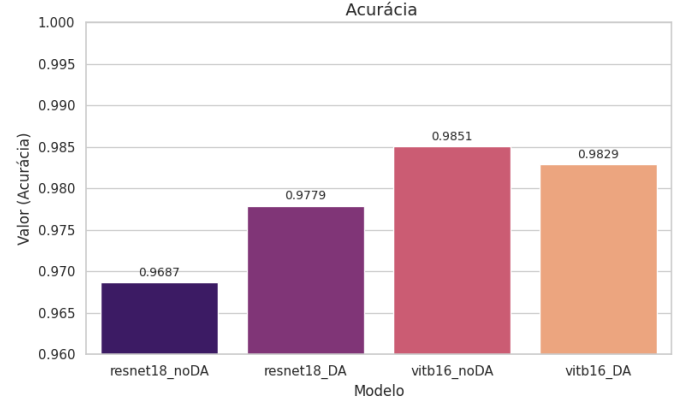


Fig. 2: Comparação das acurácias obtidas pelos modelos ResNet-18 e ViT-B/16 nos cenários com e sem data augmentation.

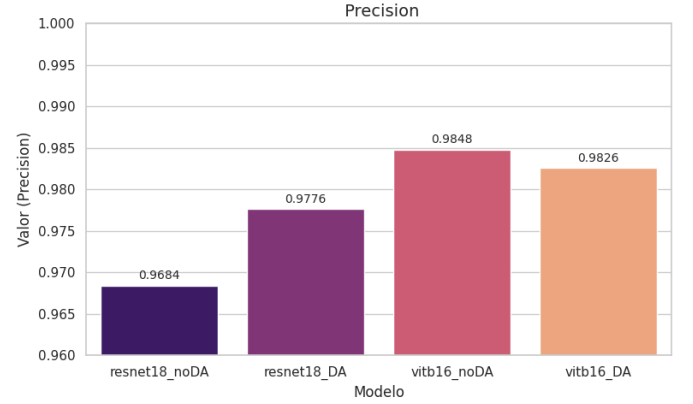


Fig. 3: Valores de precision alcançados pelos modelos avaliados, destacando o impacto do uso de data augmentation.

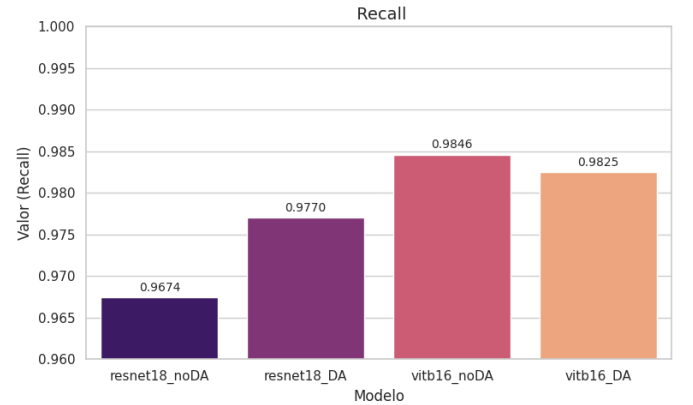


Fig. 4: Comparação dos valores de recall entre os modelos ResNet-18 e ViT-B/16 considerando diferentes estratégias de treinamento.

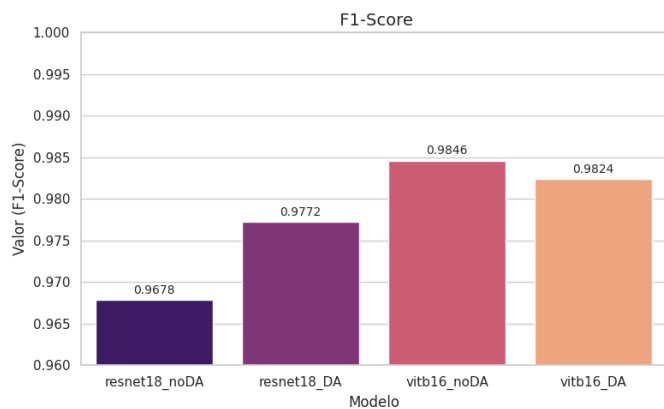


Fig. 5: Resultados de F1-score para cada modelo, permitindo analisar o equilíbrio entre precisão e sensibilidade nas classificações.

de vision transformers tendem a performar muito bem para esta tarefa. O resnet18 obteve bons resultados, no entanto foi inferior ao vit em ambos cenários, com a diferença de que foi positivamente impactado pelo aumento de dados.

B. Detalhes da Classificação

Para compreender de forma mais aprofundada o comportamento dos modelos na tarefa de classificação das quatro categorias do dataset — glioma, meningioma, pituitário e sem tumor — foram analisadas as matrizes de confusão correspondentes a cada arquitetura e a cada cenário de data augmentation. Enquanto as métricas globais fornecem uma visão geral do desempenho, as matrizes de confusão permitem identificar padrões específicos de acertos e erros, revelando quais classes apresentam maior dificuldade de discriminação. As mesmas estão nas figuras 6, 7, 8 e 9:

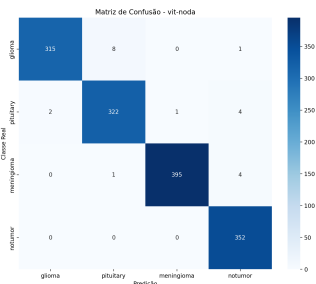


Fig. 6: Matriz de confusão obtida com o vit sem o uso de aumento de dados.

As matrizes de confusão mostram um particular padrão na classificação da classe glioma como pituitary, um erro que se repetiu tanto para o resnet18 quanto para o vitb16. Um outro erro que parece se destacar é a classificação de pituitary como notumor, no entanto esse erro foi pouco presente na classificação pelo vitb16 sem o aumento de dados. Esses padrões podem ser vistos também em outros trabalhos que

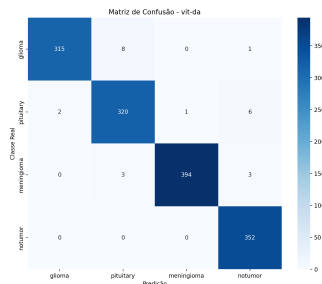


Fig. 7: Matriz de confusão para o vit na versão com aumento de dados.

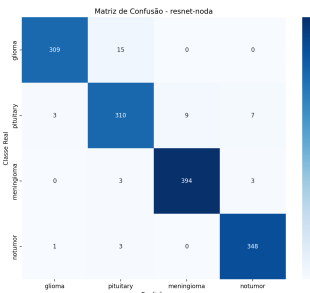


Fig. 8: Matriz de confusão da ResNet-18 sem data augmentation.

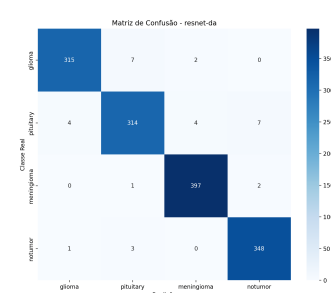


Fig. 9: Matriz de confusão da ResNet-18 com data augmentation.

avaliam a matriz de confusão como [Disci et al., 2025] e outros.

C. Impacto do Data Augmentation

O uso do data augmentation impactou de maneira distinta as duas arquiteturas. Para a ResNet-18, o uso de técnicas como flip horizontal, rotações, jitter de cor e desfoque gaussiano resultou em uma melhora nas quatro métricas, com destaque para a acurácia, que passou de 0.9687 para 0.9779. Essa influência pode ser vista no gráfico da figura 10:

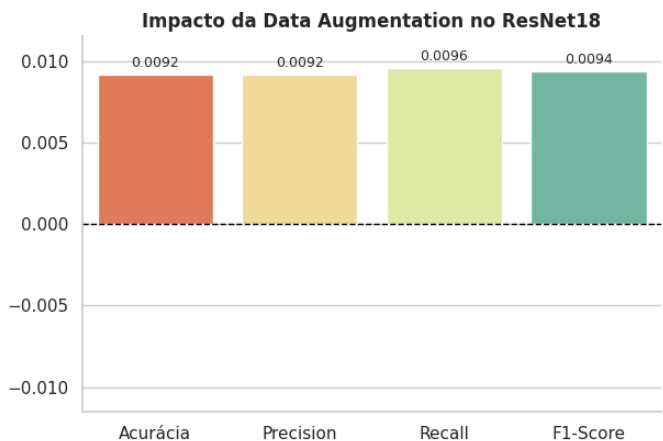


Fig. 10: Gráfico com o impacto do aumento de dados na RESNET-18, os valores acima da linha tracejada (valores da resnet sem data augmentation) indicam que houve uma ganho nas métricas.

Já para o vitb16 o mesmo aumento de dados parece ter causado um impacto negativo em todas as métricas, oque pode ser visto na figura 11:

Essa diferença no ganho de métricas pode ser explicada pelo fato de que modelos baseados em convolução tradicional tendem a se beneficiar mais de alterações como rotações e color jitter, pois essas transformações ampliam o repertório de padrões locais aprendidos pelos filtros convolucionais. Já os Vision Transformers, por seu mecanismo de auto-atenção, que permite capturar relações globais na imagem desde as

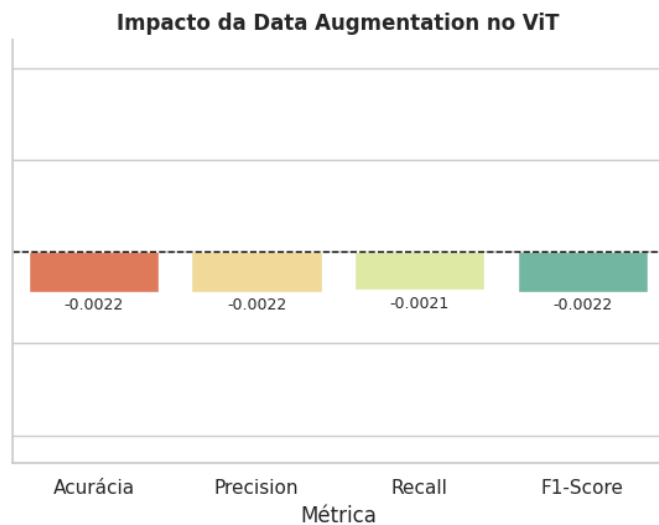


Fig. 11: Gráfico com o impacto do aumento de dados na ViT-B/16, os valores qbaixo da linha tracejada (vit sem data augmentation) indicam que houve uma perda nas métricas.

primeiras camadas, o tornam menos dependente de variações artificiais introduzidas pelo aumento de dados, assim como hipotetizado por [de Oliveira Costa et al., 2025].

Assim, os resultados confirmam um paradigma bastante presente na literatura e sugerem que o uso de data augmentation deve ser cuidadosamente ajustado ao tipo de arquitetura, podendo trazer grandes ganhos para CNNs, mas exigindo maior cautela e, possivelmente, estratégias mais específicas quando aplicado a modelos baseados em Transformers.

V. CONCLUSÃO

Neste trabalho, investigou-se o desempenho de duas arquiteturas de deep learning — ResNet-18 e ViT-B/16 — na tarefa de classificação de tumores cerebrais em imagens de ressonância magnética, comparando seus resultados em cenários com e sem data augmentation. Utilizando o dataset Brain Tumor MRI, composto por quatro classes bastante diversas, foi realizada uma avaliação baseada em métricas de acurácia, precisão, recall e F1-score.

Os resultados mostraram que ambas as arquiteturas alcançaram bons desempenhos. Entretanto, diferenças importantes foram observadas entre os modelos. A ResNet-18 apresentou ganhos quando o data augmentation foi aplicado, demonstrando melhor generalização e aumento nas métricas com o uso de transformações como random flip, rotação, jitter de cor e desfoque gaussiano. Por outro lado, o ViT-B/16 obteve seu melhor desempenho sem o uso de aumento de dados, sugerindo que o mecanismo de autoatenção já captura informações globais de maneira eficaz, reduzindo a necessidade de variabilidade artificial no treinamento. Entre todas as configurações avaliadas, o ViT-B/16 sem augmentation apresentou a maior acurácia (0.9851), assim como as outras métricas.

De modo geral, este estudo reforça a importância de alinhar as estratégias de data augmentation às características próprias de cada arquitetura. Os resultados obtidos contribuem para o entendimento mais aprofundado do comportamento de CNNs e Vision Transformers em tarefas de classificação de imagens e podem orientar pesquisas futuras e o desenvolvimento de sistemas de apoio ao diagnóstico mais eficientes e confiáveis. Os mesmos também demonstram a possibilidade de uso desses métodos clinicamente, dadas suas boas métricas.

REFERÊNCIAS BIBLIOGRÁFICAS

- [Alemayehu, 2025] Alemayehu, N. (2025). Light weight cnn for classification of brain tumors from mri images. *arXiv preprint arXiv:2504.21188*.
- [Asiri et al., 2023] Asiri, A. A., Shaf, A., Ali, T., Pasha, M. A., Aamir, M., Irfan, M., Alqahtani, S., Alghamdi, A. J., Alghamdi, A. H., Alshamrani, A. F. A., et al. (2023). Advancing brain tumor classification through fine-tuned vision transformers: A comparative study of pre-trained models. *Sensors*, 23(18):7913.
- [de Oliveira Costa et al., 2025] de Oliveira Costa, P. L., de Oliveira Costa, T. M., Moreira, L. F. R., Silva, L. H. F. P., and Mari, J. F. (2025). Classification of agricultural pests through digital images using deep learning. *Revista de Informática Teórica e Aplicada*, 32(1):18–25.
- [Disci et al., 2025] Disci, R., Gurcan, F., and Soylu, A. (2025). Advanced brain tumor classification in mr images using transfer learning and pre-trained deep cnn models. *Cancers*, 17(1):121.
- [Dosovitskiy, 2020] Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [GloboCan, 2023,] GloboCan, 2023. Global burden of brain and central nervous system cancer in 185 countries, and projections up to 2050: a population-based systematic analysis of globocan 2022. <https://pubmed.ncbi.nlm.nih.gov/40720078/>. Acesso em: 23 nov. 2025.
- [Gómez-Guzmán et al., 2023] Gómez-Guzmán, M. A., Jiménez-Beristáin, L., García-Guerrero, E. E., López-Bonilla, O. R., Tamayo-Perez, U. J., Esqueda-Elizondo, J. J., Palomino-Vizcaino, K., and Inzunza-González, E. (2023). Classifying brain tumors on magnetic resonance imaging by using convolutional neural networks. *Electronics*, 12(4):955.
- [He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- [Instituto Oncoguia, 2025,] Instituto Oncoguia, 2025. Estatísticas para tumores cerebrais. <https://www.oncoguia.org.br/conteudo/estatistica-para-tumores-cerebrais-snc/7343/293/>. Acesso em: 23 nov. 2025.
- [PMC, 2025,] PMC, 2025. The global, regional, and national brain and central nervous system cancer burden and trends from 1990 to 2021: an analysis based on the global burden of disease study 2021. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12213423>. Acesso em: 23 nov. 2025.
- [SILVA et al., 2024] SILVA, U. K. M. A. d. et al. (2024). Classificação de tumores cerebrais: um estudo comparativo entre rede neural convolucional e rede neural convolucional com mecanismo de atenção.