

Comparação de Arquiteturas Convolucionais Profundas para Classificação de Tumores Cerebrais em Imagens de Ressonância Magnética

Thamara Leandra de Deus Melo

Universidade Federal de Viçosa – Campus Rio Paranaíba

thamara.melo@ufv.br

Resumo—O presente trabalho apresenta o desenvolvimento e a comparação de duas arquiteturas de redes neurais convolucionais (CNNs) aplicadas à classificação de tumores cerebrais em imagens de ressonância magnética (MRI). Foram avaliadas as arquiteturas ResNet18 e MobileNetV3-Small, ambas pré-treinadas no ImageNet, com o objetivo de investigar o impacto da profundidade e da eficiência computacional sobre o desempenho da classificação. Utilizou-se o conjunto de dados *Kaggle Brain MRI Dataset*, submetido a técnicas de pré-processamento e aumento de dados. As redes foram treinadas no *framework* PyTorch e avaliadas por meio das métricas de acurácia, precisão, *recall* e F1-score. Os resultados mostraram que a ResNet18 obteve desempenho superior, com acurácia de 96,8%, enquanto a MobileNetV3-Small alcançou 93,6%, evidenciando o compromisso entre complexidade e desempenho das arquiteturas. Este estudo fornece uma análise comparativa clara sobre o uso de modelos convolucionais em diagnósticos médicos automatizados.

Palavras-chave — Classificação de imagens, ResNet18, MobileNetV3, Tumores cerebrais, Visão computacional.

I. INTRODUÇÃO

O câncer cerebral constitui uma das patologias mais desafiadoras da neurologia moderna, caracterizando-se pelo crescimento anormal de células no tecido cerebral, frequentemente resultante de mutações genéticas que afetam os mecanismos naturais de regulação celular [1], [2]. A complexidade dessas formações tumorais exige abordagens diagnósticas e terapêuticas cada vez mais precisas, sendo as Imagens de Ressonância Magnética (MRI) uma das principais ferramentas para a detecção não invasiva dessas anomalias.

Apesar do avanço das tecnologias de imagem, a interpretação das imagens de MRI ainda é fortemente dependente da expertise dos profissionais de saúde, o que pode levar à variabilidade nos diagnósticos e limitações em termos de escalabilidade e padronização [3]. Nesse contexto, os sistemas baseados em Inteligência Artificial (IA), em especial as Convolucional Neural Networks (CNNs), têm ganhado destaque por sua capacidade de extrair automaticamente padrões relevantes e realizar classificações com elevado grau de precisão [4].

Com o avanço do aprendizado profundo, as redes neurais convolucionais (CNNs) tornaram-se ferramentas essenciais para extração automática de padrões complexos em dados visuais. Essas redes permitem a detecção de anomalias com alto grau de precisão, reduzindo a dependência da análise manual. Entre as CNNs mais utilizadas estão as arquiteturas

ResNet e MobileNet, conhecidas respectivamente por sua profundidade e eficiência computacional.

Este trabalho tem como objetivo comparar o desempenho das arquiteturas ResNet18 e MobileNetV3-Small na classificação de tumores cerebrais em imagens de MRI, implementadas com o *framework* PyTorch e utilizando *transfer learning*. O estudo enfatiza o equilíbrio entre acurácia e custo computacional, visando contribuir para o desenvolvimento de soluções práticas e eficientes na área de diagnóstico assistido por computador.

II. REVISÃO BIBLIOGRÁFICA

A detecção e a classificação de tumores cerebrais em imagens de MRI têm sido objeto de ampla investigação, dada a gravidade dessa condição e a necessidade de diagnósticos precoces e precisos. Técnicas de aprendizado de máquina e aprendizado profundo têm sido amplamente aplicadas nesse contexto, especialmente as CNNs, que se mostraram eficientes na extração de padrões discriminativos em imagens médicas [5], [6].

Shah et al. [7] apresentaram um modelo baseado na arquitetura EfficientNet-B0 para a classificação de tumores cerebrais, combinado com técnicas de *fine-tuning* e aumento de dados. Os autores compararam o modelo com arquiteturas tradicionais, como VGG16, InceptionV3, Xception, ResNet50 e InceptionResNetV2. O modelo proposto superou todas as bases de comparação, alcançando uma acurácia de 98,87% e uma AUC de 0,988.

Khaliki and Basarslan [8] compararam o desempenho de CNNs convencionais e de modelos pré-treinados, como VGG19, EfficientNetB4 e InceptionV3, em imagens do conjunto de dados do Kaggle. Os resultados demonstraram que o uso de *transfer learning* melhora significativamente a classificação, especialmente em cenários com conjuntos de dados limitados.

Jing et al. [9] propuseram um método de fusão multimodal de imagens médicas denominado AMMNet, que integra mecanismos de atenção e múltiplas arquiteturas convolucionais com o objetivo de melhorar a nitidez e a preservação de bordas em imagens fundidas. O modelo combina convoluções multiescala para extração de características superficiais com

uma DenseNet aprimorada para extração de características profundas, além de um mecanismo de atenção híbrido composto por módulos de atenção de canal eficiente e atenção espacial. Para evitar perda de textura, os autores também desenvolveram uma rede de decomposição formada por um módulo CNN convencional acoplado a uma MobileNetV3, responsável por recuperar detalhes finos da imagem fundida ao compará-la com as imagens originais.

Em contraste com os estudos citados, que investigam arquiteturas profundas ou estratégias de fusão multimodal, o presente trabalho dedica-se exclusivamente à classificação de tumores cerebrais em imagens de MRI, utilizando apenas imagens em escala de cinza e modelos pré-treinados, sem recorrer à fusão de modalidades ou a redes excessivamente complexas. Esta pesquisa compara duas redes com níveis distintos de complexidade computacional, ResNet18 e MobileNetV3-Small, com o objetivo de analisar a relação entre desempenho e custo de processamento. Os resultados obtidos indicam que modelos leves podem alcançar desempenho satisfatório em aplicações clínicas, enquanto arquiteturas mais profundas garantem maior precisão, fornecendo subsídios objetivos para a escolha de modelos em sistemas de diagnóstico assistido por computador.

III. MATERIAIS E MÉTODOS

O conjunto de dados utilizado neste estudo foi o *Kaggle Brain MRI Dataset*, amplamente empregado em pesquisas de visão computacional aplicadas à detecção de tumores cerebrais. O dataset contém quatro classes de imagens de ressonância magnética (glioma, meningioma, pituitária e sem tumor), totalizando 5.442 amostras após a etapa de limpeza, como representado na Tabela I. Como o conjunto não apresentava divisão prévia, foi adotada a validação cruzada do tipo *hold-out*, com separação em conjuntos de treinamento, validação e teste. As imagens foram redimensionadas para 224×224 pixels, convertidas para escala de cinza e normalizadas no intervalo $[0, 1]$.

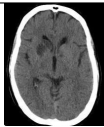
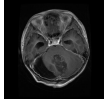
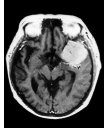
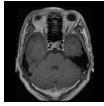
A. Pré-processamento e Aumento de Dados

O pré-processamento incluiu normalização dos níveis de intensidade e padronização dimensional. Para aumentar a robustez dos modelos, foram aplicadas estratégias de *data augmentation*, consistindo em rotações aleatórias, inversão horizontal, variação de brilho e ajuste de contraste, reduzindo o risco de sobreajuste.

B. Modelos e Arquiteturas

ResNet-18: A ResNet-18 integra a família de redes residuais proposta por He et al. [10], concebida para permitir o treinamento de arquiteturas profundas sem degradação de desempenho. Diferentemente das CNNs convencionais, esse modelo introduz conexões de atalho que possibilitam a aprendizagem de funções residuais, facilitando o fluxo de gradiente ao longo das camadas e reduzindo problemas como a explosão ou o desaparecimento do gradiente. Composta por 18 camadas treináveis, a arquitetura oferece um equilíbrio entre profundidade e custo computacional, sendo amplamente

Tabela I
DISTRIBUIÇÃO DAS CLASSES COM EXEMPLOS VISUAIS.

Classe	Quantidade	Exemplo
No Tumor	1.350	
Glioma	1.321	
Meningioma	1.331	
Pituitário	1.440	
Total	7023	

utilizada como modelo base em tarefas de classificação de imagens. Conforme discutido por Goodfellow et al. [4], o uso de blocos residuais representa um avanço significativo no aprendizado profundo, ao permitir redes mais estáveis e eficientes.

No presente estudo, a ResNet-18 foi utilizada com pesos pré-treinados (*ImageNet*) e teve sua camada totalmente conectada final substituída por uma nova camada com quatro neurônios, correspondentes às classes do conjunto de dados de tumores cerebrais. A Figura 1 ilustra a organização dos blocos residuais que compõem o modelo, destacando como as conexões de atalho contribuem para a propagação eficiente do gradiente durante o treinamento.

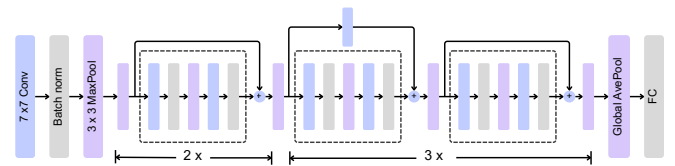


Figura 1. Schematic of the ResNet 18 architecture with basic residual blocks and four stages of increasing channel width.

MobileNetV3-Small: A MobileNetV3-Small integra a terceira geração da família MobileNet, projetada para operar em dispositivos com restrições computacionais, como sistemas embarcados e plataformas móveis [11]. Sua arquitetura combina convoluções separáveis em profundidade, blocos inversos com gargalo e a função de ativação *h-swish*, o que possibilita reduzir significativamente o número de parâmetros sem comprometer o desempenho. A variante *Small* foi otimizada para cenários de inferência em tempo real, oferecendo eficiência

superior em comparação com redes mais profundas, como ResNet e DenseNet.

Diferentemente de arquiteturas tradicionais que priorizam exclusivamente a acurácia, a MobileNetV3-Small foi projetada com base em técnicas de busca neural e otimização orientada a hardware, incorporando módulos Squeeze-and-Excitation (SE) (*Squeeze-and-Excitation*) responsáveis por recalibrar dinamicamente os canais de ativação [12]. Essas características tornam o modelo particularmente adequado para aplicações médicas que exigem baixo custo computacional aliado a boa capacidade de generalização.

Neste estudo, a MobileNetV3-Small foi empregada com pesos pré-treinados no *ImageNet*, sendo sua camada totalmente conectada final substituída por uma nova camada com quatro neurônios, correspondentes às classes do conjunto de dados de tumores cerebrais avaliadas.

C. Configurações de Treinamento

Os experimentos foram conduzidos em GPU utilizando o *framework* PyTorch. Ambos os modelos foram treinados por 20 épocas com *batch size* de 32, otimizador Adam e taxa de aprendizado inicial de 0,001. A função de perda utilizada foi a *Cross-Entropy Loss*. Durante o treinamento, foram monitoradas a perda e a acurácia no conjunto de validação para avaliação de convergência.

D. Métricas de Avaliação

A avaliação dos modelos foi realizada utilizando métricas amplamente empregadas em tarefas de classificação supervisionada, a fim de mensurar diferentes aspectos de desempenho. A partir da matriz de confusão, consideram-se: *TP* (verdadeiros positivos), *TN* (verdadeiros negativos), *FP* (falsos positivos) e *FN* (falsos negativos). As métricas adotadas neste estudo são definidas formalmente como segue.

- **Acurácia** (*Accuracy*):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Representa a proporção total de instâncias corretamente classificadas, fornecendo uma medida global de desempenho do modelo [4].

- **Precisão** (*Precision*):

$$Precision = \frac{TP}{TP + FP}$$

Quantifica o percentual de previsões positivas que são de fato corretas, sendo relevante em cenários nos quais falsos positivos devem ser minimizados [6].

- **Sensibilidade** (*Recall*):

$$Recall = \frac{TP}{TP + FN}$$

Mede a capacidade do modelo em identificar corretamente as instâncias pertencentes à classe positiva, ou seja, o quanto o modelo deixa de detectar [6].

- **F1-Score**:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Corresponde à média harmônica entre Precisão e Recall, sendo particularmente útil quando existe desequilíbrio entre classes [5].

Todas as métricas foram calculadas utilizando a biblioteca *scikit-learn*, com valores reportados individualmente para cada classe bem como em médias ponderadas (*weighted*), conforme apresentado nas tabelas de resultados.

IV. RESULTADOS E DISCUSSÃO

A Tabela II apresenta as métricas de avaliação obtidas pelos dois modelos após o treinamento completo. A ResNet18 apresentou desempenho superior em todas as métricas analisadas, alcançando acurácia global de 96,79% e F1-score médio de 96,59%. A MobileNetV3-Small, embora inferior, obteve resultados consistentes, com acurácia de 93,59% e F1-score de 93,54%, demonstrando boa capacidade de generalização mesmo com número reduzido de parâmetros.

Tabela II
RESULTADOS COMPARATIVOS ENTRE AS ARQUITETURAS AVALIADAS

Modelo	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)
ResNet18	96.79	97.04	96.79	96.81
MobileNetV3-Small	93.59	94.18	93.59	93.54

A análise das métricas por classe confirma o comportamento observado. A ResNet18 apresentou desempenho equilibrado em todas as categorias, com destaque para a classe *notumor*, que obteve F1-score de 99,38%, e *pituitária*, com F1-score de 96,57%. Já a MobileNetV3-Small apresentou maior queda de desempenho nas classes *pituitária* e *meningioma*, com F1-score de 90,14% e 93,02%, respectivamente.

Esses resultados sugerem que a ResNet18 consegue extrair padrões mais discriminativos entre tipos de tumores, enquanto a MobileNetV3-Small apresenta limitações principalmente em classes com maior sobreposição visual.

A análise das matrizes de confusão das Figuras 2 e 3 permite compreender com maior precisão o comportamento das arquiteturas em cada classe. A ResNet18 apresentou desempenho extremamente estável, classificando corretamente todos os casos de *glioma* (300 acertos) e quase a totalidade de imagens *notumor* (404 acertos em 405 amostras). Os erros concentraram-se majoritariamente na classe *meningioma*, com 18 imagens incorretamente rotuladas como *glioma*, e na classe *pituitary*, com 16 amostras classificadas como *glioma*. Ainda assim, a taxa de erro global permaneceu reduzida, reforçando a capacidade da rede em discriminar padrões entre tumores estruturalmente semelhantes.

No caso da MobileNetV3, observou-se aumento do número de erros entre classes visualmente próximas. A classe *glioma* apresentou 277 acertos e 21 amostras confundidas com *meningioma*, enquanto a classe *pituitary* registrou maior dispersão, com 37 casos incorretamente atribuídos a *notumor*. Apesar disso, o modelo manteve desempenho elevado nas demais classes, com 300 acertos em *meningioma* e 403 em *notumor*, evidenciando que a maior parte dos equívocos ocorre

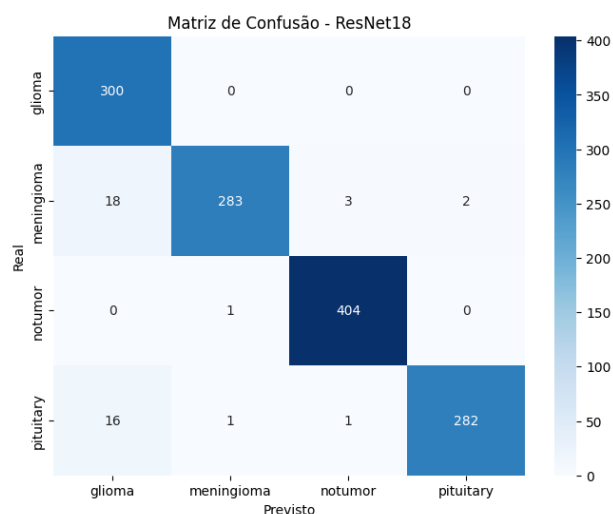


Figura 2. Matriz de confusão obtida para a arquitetura ResNet18

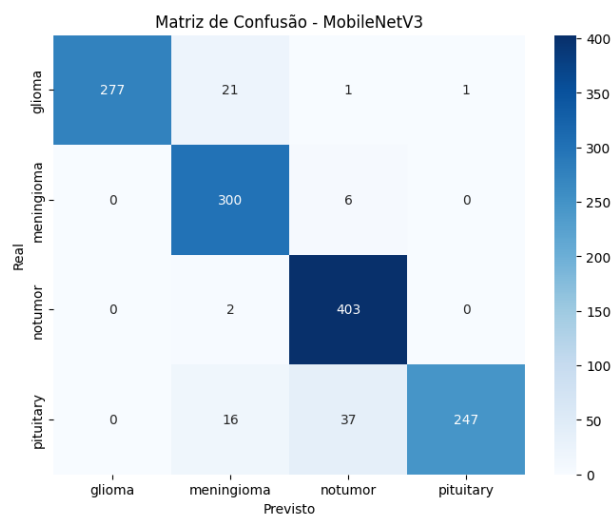


Figura 3. Matriz de confusão obtida para a arquitetura MobileNetV3.

em categorias que compartilham características morfológicas similares.

De forma geral, os resultados confirmam a superioridade da ResNet18 na separação das quatro classes avaliadas, tanto em métricas quantitativas quanto em consistência visual da matriz de confusão. A MobileNetV3-Small, embora apresente maior taxa de confusão entre tumores, mantém desempenho competitivo considerando sua arquitetura compacta, o que reforça seu potencial para aplicações em dispositivos com restrições de processamento.

V. CONCLUSÃO

Este estudo realizou uma comparação entre as arquiteturas ResNet18 e MobileNetV3 aplicadas à classificação de tumores cerebrais em imagens de ressonância magnética. Os resultados demonstraram que a ResNet18 apresentou o melhor desempenho geral, alcançando acurácia de 96,79% e F1-score médio

de 96,59%. A MobileNetV3-Small, embora inferior, obteve desempenho competitivo com acurácia de 93,59%, destacando-se pela eficiência em termos de número de parâmetros e velocidade de inferência.

Os achados indicam que redes mais profundas, como a ResNet18, são mais adequadas para ambientes clínicos que demandam máxima precisão diagnóstica. Por outro lado, arquiteturas compactas como a MobileNetV3-Small apresentam potencial para uso em sistemas portáteis, aplicações embarcadas e dispositivos com restrições computacionais, mantendo desempenho satisfatório.

Como trabalhos futuros, sugere-se a aplicação de técnicas de otimização de hiperparâmetros, utilização de estratégias de *fine-tuning* mais agressivas, teste de arquiteturas híbridas e investigação de métodos de calibração de confiança para aumentar a interpretabilidade dos modelos em ambientes médicos reais.

REFERÊNCIAS

- [1] R. Stupp, W. P. Mason, M. J. van den Bent, M. Weller, B. Fisher, M. J. B. Taphoorn, K. Belanger, A. A. Brandes, C. Marosi, U. Bogdahn, J. Curschmann, R. Janzer, S. Ludwin, T. Gorlia, A. Allgeier, J.-P. Lacombe, J. Cairncross, E. Shaw, R. K. Jain, M. J. Mehta, J. G. G. Rampling, G. J. Wick, W. Wick, M. Westphal, J. C. Tonn, J. P. Mirimanoff, E. Eisenhauer, and M. E. Hegi, "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma," *New England Journal of Medicine*, vol. 352, no. 10, pp. 987–996, 2005.
- [2] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and E. D. W., "The 2021 who classification of tumors of the central nervous system: a summary," *Neuro-Oncology*, vol. 23, no. 8, pp. 1231–1251, 2021.
- [3] P. Y. Wen and M. Butler, "Glioblastoma multiforme: biology and management," in *Abeloff's Clinical Oncology*, 6th ed. Philadelphia: Elsevier, 2020, ch. 77.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [5] S. Solanki, U. P. Singh, S. S. Chouhan, and S. Jain, "Brain tumor detection and classification using intelligence techniques: An overview," *IEEE Access*, vol. 11, pp. 12 870–12 886, 2023.
- [6] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "Brain tumor detection and classification using machine learning: A comprehensive survey," *Visual Computing for Industry, Biomedicine, and Art*, vol. 4, no. 1, pp. 1–19, 2022.
- [7] H. A. Shah, F. Saeed, S. Yun, J.-H. Park, A. Paul, and J.-M. Kang, "A robust approach for brain tumor detection in magnetic resonance images using finetuned efficientnet," *IEEE Access*, vol. 10, pp. 65 426–65 438, 2022.
- [8] M. Khaliki and M. Başarslan, "Brain tumor detection from images and comparison with transfer learning methods and 3-layer cnn," *Scientific Reports*, vol. 14, no. 1, p. 2664, February 2024.
- [9] J. Di, W. Guo, J. Liu, L. Ren, and J. Lian, "Ammnet: A multimodal medical image fusion method based on an attention mechanism and mobilenetv3," *Biomedical Signal Processing and Control*, vol. 96, p. 106561, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809424006190>
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [11] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for mobilenetv3," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019, pp. 1314–1324.
- [12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.