

Classificação de Tumores Cerebrais em Imagens de Ressonância Magnética

Andressa G. Moreira¹, Stefane A. dos Santos¹, Michele F. de Oliveira¹,
Iális C. de Paula Júnior¹, Débora F. de Assis²

¹Programa de Pós-Graduação em Engenharia Elétrica e da Computação –
Universidade Federal do Ceará (UFC) – Sobral – CE – Brasil ,

²Laboratório de Engenharia de Sistemas de Computação –
Universidade Federal do Ceará (UFC) – Fortaleza – CE – Brasil

Abstract. *Brain cancer is characterized by the development of malignant tumors. Therefore, early detection is crucial for patient survival. Advances in Artificial Intelligence (AI) have improved medical image analysis. However, classifying projects is still a challenging task. In this study, the transfer learning technique is used to classify the types of tumors into Meningioma, Glioma, Pituitary lesions and cases without tumor, based on magnetic resonance images. For this, the AlexNet, DenseNet201, EfficientNetB7, MobileNetV2 and ResNet50 architectures are used. The EfficientNetB7 network obtained the most promising results, reaching 97.68% for test accuracy.*

Resumo. *O câncer de cérebro é caracterizado pelo desenvolvimento de tumores cerebrais malignos. Dessa forma, a detecção precoce é crucial para a sobrevivência dos pacientes. Os avanços em Inteligência Artificial (IA) têm aprimorado a análise de imagens médicas. Entretanto, a classificação de tumores cerebrais ainda é uma tarefa desafiadora. Neste estudo, é utilizada a técnica de transfer learning para classificar os tipos de tumores cerebrais em Meningioma, Glioma, Hipofisário e casos sem tumor, a partir de imagens de ressonância magnética. Para isso, são empregadas as arquiteturas AlexNet, DenseNet201, EfficientNetB7, MobileNetV2 e ResNet50. A rede EfficientNetB7 obteve os resultados mais promissores, alcançando 97,68% para a acurácia de teste.*

1. Introdução

O cérebro é um dos órgãos mais complexo e sensível do corpo humano. Segundo o Instituto Nacional de Câncer (INCA), no Brasil são registrados cerca de 11 mil novos casos de pacientes com doenças tumorais cerebrais a cada ano, com uma taxa de mortalidade de aproximadamente 84% [Instituto Nacional de Câncer (INCA) 2022]. Para realizar o diagnóstico de tumores cerebrais, os exames de imagens como Tomografia Computadorizada (TC) e Ressonância Magnética (RM) são as principais tecnologias utilizadas. Vale ressaltar que a ressonância magnética é a técnica de imagem médica mais comumente usada por se tratar de um processo não invasivo [Raj et al. 2024].

O diagnóstico precoce de tumores cerebrais pode aumentar as taxas de sobrevivência dos pacientes e fornecer opções oportunas para os planos de tratamento adequados. Entretanto, a classificação de tumores cerebrais é uma tarefa desafiadora por diversos fatores, por exemplo, alterações no tamanho, formato e posição do tumor no cérebro, dependendo do paciente. Além disso, a classificação de características tumorais e normais é subjetiva e pode incluir uma margem de erro significativa [Ottom et al. 2022].

Nesse contexto, abordagens baseadas em *Deep Learning (DL)* têm aumentado o sucesso em pesquisas com imagens médicas e auxiliado a classificação precoce de tumores cerebrais. Técnicas computacionais auxiliam no desenvolvimento de modelos interpretáveis, capazes de reconhecer padrões automaticamente e diagnosticar novos casos. Em geral, os sistemas automáticos de diagnóstico são destinados a auxiliar o diagnóstico médico [Padmapriya and Devi 2024]. Todavia, esses sistemas exigem uma grande quantidade de dados [Asif et al. 2022]. Logo, é necessário um esforço para desenvolver um sistema automatizado eficiente e que forneça resultados precisos.

Portanto, o objetivo desse projeto é desenvolver um fluxo de trabalho e eficaz para auxílio ao diagnóstico e para classificar tumores cerebrais, por meio de imagens de ressonância magnética, com maior precisão e performance. É proposto a utilização de arquiteturas de redes neurais que garantam a assertividade. Vale ressaltar ainda que, apesar dos inúmeros benefícios, a proposta não substitui o diagnóstico feito por um médico especialista. O fluxo de trabalho serve como uma ferramenta de apoio para auxiliar os profissionais de saúde a realizarem diagnósticos precoce com maior precisão.

Para atingir os objetivos, foram utilizados bancos de dados disponíveis publicamente na literatura de imagens de ressonância magnética. A priori, para realizar a classificação dos tipos de tumores cerebrais em Meningioma, Glioma, Hipofisário e casos sem tumor, foram utilizados os modelos pré-treinados AlexNet, DenseNet201, EfficientNetB7, MobileNetV2 e ResNet50. Por fim, o desempenho dos modelos foi avaliado pelas métricas de avaliação, como, acurácia, precisão, *recall*, *F1-Score* e especificidade. Além disso, foram realizados testes estatísticos para analisar a significância dos resultados.

Logo, as principais contribuições deste trabalho são listadas a seguir:

- Uma análise estatística abrangente é conduzida em várias arquiteturas de redes neurais para classificar os tipos de tumores cerebrais, incluindo Meningioma, Glioma, Hipofisário e casos sem tumor;
- Experimentos são conduzidos utilizando uma combinação de conjuntos de dados, visando assegurar uma generalização mais eficaz do modelo, com a inclusão de um conjunto de validação externo.

O restante do trabalho está organizado da seguinte forma: A Seção 2 apresenta uma revisão de trabalhos relacionados na literatura. Em seguida, a Seção 3 descreve as etapas metodológicas realizadas na elaboração deste trabalho. A Seção 4 apresenta uma análise e discussão dos resultados experimentais. Por fim, a Seção 5 apresenta as considerações finais deste projeto e determina os trabalhos futuros.

2. Trabalhos Relacionados

Devido à necessidade de um diagnóstico preciso e precoce e aos fatores que dificultam a atividade de detecção e classificação de tumores cerebrais, diversos estudos no âmbito da Inteligência Artificial (IA) foram realizados para auxiliar o diagnóstico automático de novos casos. Nesta seção, são abordados trabalhos da literatura que empregam o conceito de IA para a classificação de tumores cerebrais.

Os autores em El-Assiouti *et al.* (2023) [El-Assiouti et al. 2023] apresentam novas técnicas de aumento de dados, definidas como RegionInpaint e aumento do RegionMix, a fim de melhorar o desempenho da identificação de tumores cerebrais. Foram

utilizados os conjuntos de dados SPMRI e Br35H, compostos por imagens de ressonância magnética cerebral distribuídas nas classes “Tumor” e “Não tumoral”. Para a tarefa de classificação binária foram realizados testes com diversas arquiteturas pré-treinadas. O melhor resultado para a precisão de teste alcançado foi de 96,88% para o modelo VGG19 treinado com imagens do conjunto de dados Br35H com técnicas de aumento.

Em Bindu e Devi (2024) [Bindu and Devi 2024], o objetivo principal reside na classificação binária de tumores cerebrais. Para este propósito, o estudo empregou o método de *transfer learning* para extrair os recursos de imagens de ressonância magnética, utilizando quatro Redes Neurais Convolucionais (CNNs) pré-treinadas. Adicionalmente, os autores consideraram imagens segmentadas do conjunto de dados original para treinamento. Os resultados experimentais demonstram uma classificação de 96,98%, evidenciando a eficácia do método de *transfer learning* empregado.

Portanto, em diversas aplicações médicas, modelos de aprendizagem profunda são utilizados em tarefas de auxílio ao diagnóstico. Alguns trabalhos se concentram no desenvolvimento de modelos para a classificação de tumores cerebrais [Srivastava et al. 2023, Islam et al. 2023, Patare et al. 2024]. Neste trabalho, é realizado um fluxo de trabalho para classificar tumores cerebrais de acordo com o tipo, auxiliando o profissional da saúde na detecção e diagnóstico da doença.

3. Materiais e Métodos

Nesta seção, está concentrada a metodologia utilizada para a elaboração deste projeto. Serão abordados detalhes sobre o fluxo dos experimentos, os conjuntos de dados utilizados, pré-processamento e aumento de dados, técnica de *transfer learning*, métricas de avaliação e os testes estatísticos utilizados durante o desenvolvimento.

3.1. Abordagens dos Experimentos

Este projeto consiste na classificação de diferentes tipos de tumores cerebrais, incluindo Meningioma, Glioma, Hipofisário e casos sem tumor. A Figura 1 apresenta as etapas realizadas no fluxo de trabalho proposto. A priori, foram aplicadas técnicas de pré-processamento nas imagens de entrada e realizado o aumento de dados do conjunto de imagens de treinamento. Em seguida, realizou-se o treinamento de arquiteturas de redes neurais pré-treinadas. Ao final dessa etapa, foi realizada a análise dos resultados, por meio de métricas de avaliação e testes estatísticos. Por fim, para avaliar o desempenho e robustez dos modelos, foi realizada uma validação externa com novos conjuntos de dados.

3.2. Base de Dados

Para o treinamento das arquiteturas foi utilizado um conjunto de dados de imagens de ressonância magnética disponível publicamente na literatura [Bhuvaji et al. 2020]. A base de dados é formada por 3264 imagens de RM com contraste aprimorado. As imagens são classificadas de acordo com os tipos de tumor cerebral: Meningioma (937 imagens), Glioma (926 imagens), Hipofisário (901 imagens) e sem tumor (500 imagens). A Figura 2 apresenta as imagens relacionadas às diferentes classes que compõem a base de dados.

Para a validação externa, foi empregada a combinação de dois conjuntos de dados públicos. O banco de dados Figshare [Cheng 2017], possui 3064 imagens de RM com

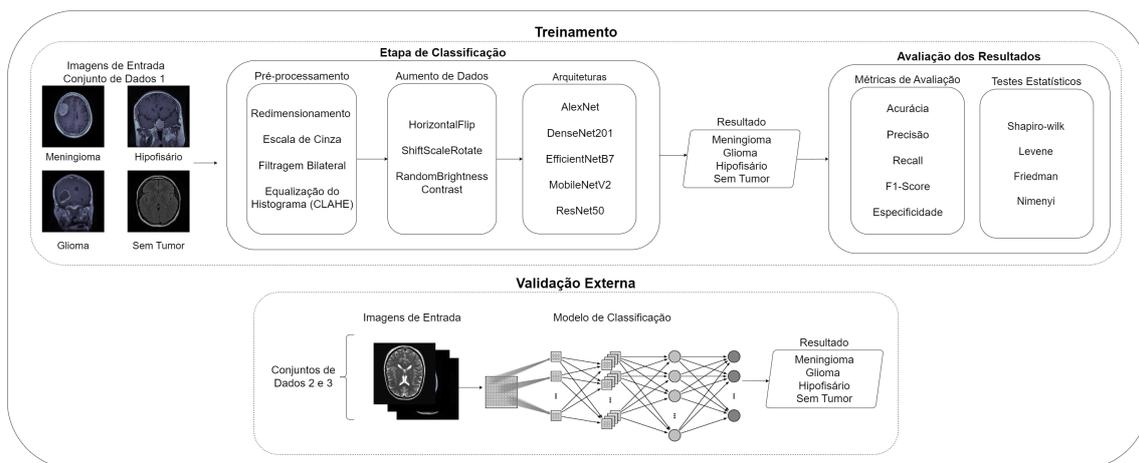


Figura 1. Fluxograma de Trabalho Proposto

contraste, obtidas de 233 pacientes e classificadas em três classes: Meningioma (708 imagens), Glioma (1426 imagens) e Hipofisário (930 imagens). Além disso, foi utilizado o banco de dados Br35H [Chakrabarty 2017], que inclui imagens de RM divididas em duas classes, com tumor e sem tumor. Para esse trabalho, foram selecionadas 1500 imagens da classe sem tumor para compor a validação externa.

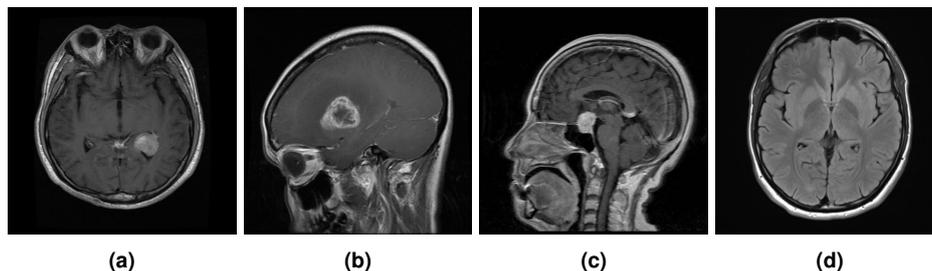


Figura 2. Amostras da Base de Dados [Bhuvaji et al. 2020]. (a) Meningioma. (b) Glioma. (c) Hipofisário. (d) Sem Tumor.

3.3. Pré-processamento

O pré-processamento de imagens consiste em transformar os dados de imagem brutos em dados de imagem limpos, com objetivo de aprimorar a qualidade da imagem e reduzir artefatos indesejados [Bow 2002]. Neste projeto, as imagens provenientes dos bancos de dados foram convertidas para escala de cinza, redimensionadas para o tamanho 224 x 224 pixels e salvas no formato PNG. Em seguida, foi aplicada a filtragem bilateral, a fim de suavizar a imagem, reduzir o ruído indesejado e preservar a nitidez das bordas [Tomasi and Manduchi 1998]. Por fim, para melhorar o contraste das imagens, foi realizada a equalização de histograma adaptativo limitado por contraste (CLAHE) [Zuiderveld 1994].

3.4. Aumento de Dados

Para melhorar a capacidade de generalização dos modelos, expandir a variabilidade do banco de dados e mitigar problemas de *overfitting*, foram utilizadas técnicas de

transformações em imagens [Chlap et al. 2021]. Vale ressaltar que as transformações não foram aplicadas para os conjuntos de validação e teste. Portanto, foram realizadas transformações afins nas imagens de treinamento, como, mudanças no brilho e contraste, inversão horizontal, mudança de escala e rotação.

3.5. *Transfer Learning*

No processo de classificação, um modelo é responsável por definir a probabilidade de uma saída pertencer à mesma classe de uma determinada entrada. Entretanto, o treinamento de uma arquitetura CNN do zero é um processo extremamente demorado. Para mitigar essa questão, é comum a aplicação da técnica de *Transfer Learning*. Nessa abordagem, são utilizados modelos CNN pré-treinados em conjuntos de dados, aplicando o conhecimento já adquirido [Alzubaidi et al. 2021, Kaya and Gürsoy 2023]. Logo, para realizar a classificação dos tipos de tumores, foram utilizadas as arquiteturas AlexNet [Krizhevsky et al. 2012], DenseNet201 [Huang et al. 2017], EfficientNetB7 [Tan and Le 2019], MobileNetV2 [Sandler et al. 2018] e ResNet50 [He et al. 2016], com pesos pré-treinados do ImageNet.

3.6. *Grid Search*

A classificação de tipos de tumores é uma tarefa desafiadora por diversos aspectos, como, o tamanho, formato e a posição do tumor. Dessa forma, para otimizar o desempenho dos modelos de classificação, utilizou-se a técnica *Grid Search*. Essa abordagem consiste em definir os melhores conjuntos de hiperparâmetros para cada arquitetura utilizada, com objetivo de maximizar a eficácia dos modelos. Nesse contexto, definiu-se os valores para os hiperparâmetros: taxa de aprendizagem, otimizador, tamanho do lote (*batch size*) e decaimento de peso (*weight decay*).

Em suma, a taxa de aprendizagem garante que o modelo atinja a convergência de maneira eficaz. O otimizador é responsável por ajustar os pesos do modelo durante o treinamento. O *batch size* define a quantidade de exemplos de treinamento usados em uma iteração. Já o *weight decay* adiciona uma penalidade à função de custo do modelo, a fim de evitar o *overfitting*. Além disso, foram aplicadas 50 épocas para cada modelo, o que representa o número de iterações durante o treinamento. A Tabela 1 apresenta os valores determinados para cada hiperparâmetro na construção dos modelos.

Arquiteturas	Taxa de Aprendizagem	Otimizador	Batch Size	Weight Decay
AlexNet	0,001	Adam	32	0,00001
DenseNet201	0,0001	Adam	32	0,0001
EfficientNetB7	0,0005	Adam	32	0,0001
MobileNetV2	0,001	Adam	32	0,0001
ResNet50	0,0001	RMSprop	32	0,00001

Tabela 1. Hiperparâmetros Otimizados com *Grid Search*

3.7. *Validação Cruzada*

Neste procedimento foi utilizada a validação cruzada *k-fold*. Nessa abordagem, o conjunto de dados é particionado aleatoriamente em $k = 10$ *folds*. Para isso, os dados foram divididos em 80% a 20% para treinamento e teste, respectivamente, com 10% dos dados

de treinamento alocados para validação. Por fim, o desempenho do modelo é determinado pela média do desempenho para todos os *folds*.

3.8. Métricas de Avaliação

No desenvolvimento de algoritmos de *Deep Learning* é imprescindível a utilização de métricas de avaliação adequadas para avaliar o desempenho dos modelos. A qualidade do treinamento é definida pelo resultado das métricas, logo, é crucial a escolha de métricas que avaliem corretamente se o modelo está de fato atendendo aos requisitos necessários.

Em tarefas de classificação, o valor verdadeiro positivo (VP) refere-se aos casos corretamente identificados como pertencentes a uma determinada categoria pelo modelo, ou seja, são os tumores de uma determinada classe que o modelo identificou corretamente. Os falsos positivos (FP) são os casos erroneamente identificados como pertencentes a uma determinada classe. Os verdadeiros negativos (VN) indicam os casos identificados corretamente como não pertencentes a uma categoria em análise. Finalmente, os falsos negativos (FN) representam os casos erroneamente identificados como não pertencendo à categoria principal.

As seguintes equações representam as métricas para avaliar os modelos de classificação: acurácia (1), precisão (2), *recall* (3), *F1-score* (4) e especificidade (5).

$$\text{Acc} = \frac{(\text{VP} + \text{VN})}{(\text{VP} + \text{VN} + \text{FP} + \text{FN})} \quad (1), \quad \text{Prec} = \frac{\text{VP}}{(\text{VP} + \text{FP})} \quad (2), \quad \text{Recall} = \frac{\text{VP}}{(\text{VP} + \text{FN})} \quad (3)$$

$$\text{F1} = \frac{2 \times (\text{Prec} \times \text{Recall})}{(\text{Prec} + \text{Recall})} \quad (4), \quad \text{Esp} = \frac{\text{VN}}{(\text{VN} + \text{FP})} \quad (5)$$

Portanto, a métrica acurácia é usada para determinar o desempenho geral do modelo, enquanto a precisão refere-se à capacidade do modelo em identificar de forma precisa os casos positivos. O *recall* indica que o modelo é capaz de identificar a maioria dos casos positivos, mesmo que também classifique erroneamente alguns casos negativos. Além disso, o *F1-score* é a média harmônica das métricas de precisão e *recall*, considerando a importância das duas as métricas. Por fim, a especificidade é uma métrica essencial para avaliar o desempenho de um modelo em problemas em que a identificação correta dos casos negativos é crucial.

3.9. Testes Estatísticos

Para compreender a significância dos resultados e identificar as possíveis diferenças entre os resultados dos modelos de classificação, foi conduzido uma análise estatística abrangente, por meio dos seguintes testes:

- **Shapiro-wilk**: Testar a suposição de normalidade das médias das métricas de avaliação para cada um dos modelos [Shapiro and Wilk 1965];
- **Levene**: Avaliar a suposição de variâncias iguais [Brown and Forsythe 1974];
- **Friedman**: Aplicados quando pelo menos uma das populações não atende ao critério de normalidade e homogeneidade. O teste não paramétrico de *Friedman* determina se existe diferença significativa entre as médias [Friedman 1937].
- **Nemenyi**: O *pós-hoc* de *Nemenyi* é responsável por identificar quais grupos são significativamente diferentes entre si [Nemenyi 1963].

3.10. Validação Externa

Para garantir a generalização e a robustez do modelo, foi realizada uma validação externa. Para isso, foram combinados dois conjuntos de dados de imagens de ressonância magnética, não utilizados durante o treinamento, validação e teste dos modelos. Desse modo, o objetivo é avaliar o desempenho dos modelos para situações distintas.

4. Resultados

Esta seção apresenta uma análise e discussão detalhada dos resultados referentes à metodologia empregada para a classificação dos tipos de tumores cerebrais em imagens de ressonância magnética. Para isso, foram comparados os resultados para as métricas de avaliação para cinco modelos pré-treinados distintos.

4.1. Classificação

Para realizar a treinamento do modelo e a classificação dos dados, a base de dados foi dividida em 80% para treinamento e 20% para teste, com 10% dos dados de treinamento alocados para validação e utilizada a validação cruzada *k-fold*. O desempenho dos modelos para a tarefa de classificação multiclasse foi analisado quantitativamente pelas métricas de acurácia de teste, precisão, *recall*, *F1-Score* e especificidade. A Tabela 2 exibe os resultados para as métricas de avaliação dos cinco modelos avaliados.

Arquitetura	Acc (%)	Prec (%)	Recall (%)	F1 (%)	Esp (%)
AlexNet	84,53	84,66	84,71	84,47	94,81
DenseNet201	97,25	97,36	97,08	97,2	99,07
EfficientNetB7	97,68	97,63	97,69	97,64	99,21
MobileNetV2	83,24	82,72	84,14	83,07	94,34
ResNet50	97,11	97,3	96,97	97,12	99,02

Tabela 2. Resultados da Classificação de Tumores Cerebrais

Os resultados obtidos sugerem que os modelos avaliados têm potencial promissor para distinguir os tumores cerebrais em Meningioma, Glioma, Hipofisário e casos sem tumor. Dentre os modelos analisados, o EfficientNetB7 obteve as melhores taxas gerais, com uma acurácia de teste de 97,68%, precisão de 97,63%, *recall* de 97,69%, *F1-Score* de 97,64% e especificidade de 99,21%. Ademais, as arquiteturas DenseNet201 e ResNet50 apresentaram desempenho semelhante em relação às métricas analisadas. Porém, as redes AlexNet e MobileNetV2 alcançaram as menores pontuações em termos métricas, com 84,53% e 83,24% de acurácia, respectivamente.

A Figura 3 exibe as matrizes de confusão para os modelos analisados. Destaca-se a matriz de confusão que resume os resultados de classificação para a rede EfficientNetB7 (Figura 3c). O classificador obteve um desempenho significativo em classificar as quatro classes, com elevado número de verdadeiros positivos e baixo valor de falsos positivos. Em suma, o modelo EfficientNetB7 classificou corretamente 314 instâncias para a classe Meningioma, 330 instâncias para a classe Glioma, 322 instâncias para a classe Hipofisário e 154 para a classe sem tumor. Em contrapartida, as redes AlexNet (Figura 3a) e MobileNetV2 (Figura 3d) apresentaram uma elevada taxa de falsos positivos, especialmente, para as classes Meningioma e Glioma.

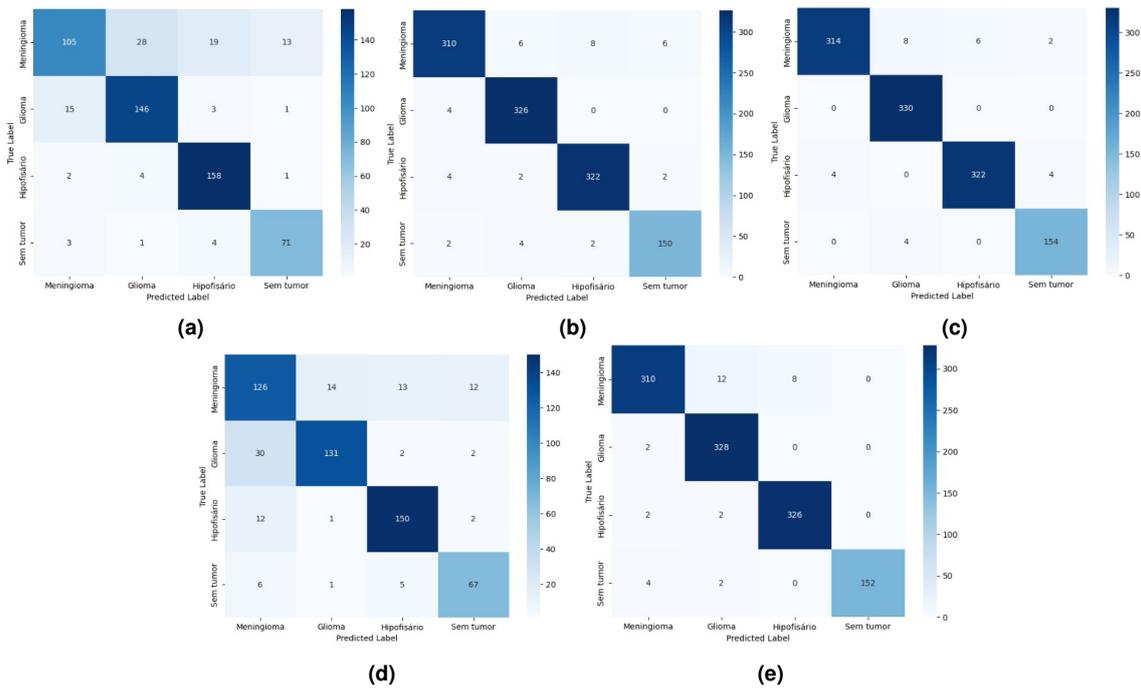


Figura 3. Matriz de Confusão. (a) AlexNet. (b) DenseNet201. (c) EfficientNetB7. (d) MobileNetV2. (e) ResNet50.

4.2. Testes Estatísticos

Para avaliar a significância dos resultados e identificar possíveis diferenças entre as médias das métricas utilizadas, foi conduzida uma análise abrangente por meio de testes estatísticos. Dessa forma, para validar as premissas de normalidade dos dados, foi realizado o teste de *Shapiro-wilk*. Em seguida, foi conduzido o teste de *Levene* para avaliar a homogeneidade entre populações. Entretanto, as hipóteses nulas de normalidade em todas as populações e de homocedasticidade dos dados foram rejeitadas. Logo, como os dados não atenderam aos requisitos de normalidade e homogeneidade, foram aplicados os testes não paramétricos.

O teste de *Friedman* foi utilizado para determinar se existem diferenças significativas entre os valores médios das populações. A Tabela 3 exibe os resultados do teste não paramétrico de Friedman, com um nível de significância de 5%. É possível observar que em todos os casos obteve-se um valor p inferior a 0,05. Isso indica que deve-se rejeitar a hipótese nula, ou seja, existem diferenças significativas entre os métodos avaliados.

	Acc	Prec	Recall	F1	Esp
Estatística de Teste	40,43	38,68	41,84	39,88	40,51
Valor p	$1,22e^{-07}$	$2,74e^{-07}$	$6,32e^{-07}$	$1,57e^{-07}$	$1,17e^{-07}$

Tabela 3. Resultados do teste de Friedman

Logo, devido à significância dos resultados obtidos no teste de Friedman, foi utilizado, em seguida, o *pós-hoc* de *Nemenyi* para inferir quais diferenças são estatisticamente relevantes. O teste foi utilizado para determinar a diferença estatística entre os pares de métodos, e, assim, avaliar o desempenho dos modelos aplicados.

A Tabela 4 apresenta a análise do teste de *Nemenyi* para a métrica acurácia, com um nível de significância de 5%. Estão destacados os resultados para qual o valor p é inferior a 0,05. Isso sugere que a hipótese nula deve ser rejeitada, ou seja, existem diferenças estatísticas entre os pares de métodos analisados. Pode-se observar que, para a métrica de acurácia, as arquiteturas AlexNet, DenseNet201 e EfficientNetB7 apresentam diferenças significativas. Entre essas redes, a AlexNet exibe a menor pontuação de acurácia. Além disso, a rede MobileNetV2 também apresenta diferenças significativas em relação às redes DenseNet201, EfficientNetB7 e ResNet50, registrando o menor valor de acurácia entre todas as arquiteturas.

Acurácia					
	AlexNet	DenseNet201	EfficientNetB7	MobileNetV2	ResNet50
AlexNet	1,0	0,02	0,001	0,96	0,05
DenseNet201	0,02	1,0	0,97	0,0008	0,99
EfficientNetB7	0,001	0,97	1,0	0,00012	0,91
MobileNetV2	0,96	0,0008	0,00012	1,0	0,002
ResNet50	0,05	0,99	0,91	0,002	1,0

Tabela 4. Resultados do teste de *Nemenyi* para a métrica Acurácia

Diante disso, das redes avaliadas, a AlexNet e a MobileNetV2 exibem o desempenho mais limitado. Assim, no contexto da classificação de tumores cerebrais, as arquiteturas DenseNet201, EfficientNetB7 e ResNet50 são as mais eficazes, visto que obtiveram diferenças insignificantes para a métrica de acurácia.

4.3. Validação Externa

Por fim, foi realizado o procedimento de validação externa para avaliar a capacidade de generalização das arquiteturas DenseNet201, EfficientNetB7 e ResNet50. A Tabela 5 apresenta os resultados dos modelos avaliados. As arquiteturas apresentaram resultados semelhantes e competitivos na validação externa para os conjuntos de dados Figshare e Br35H. A rede EfficientNetB7, apresentou as melhores pontuações para as métricas analisadas, com 99,01% de acurácia, precisão de 98,79%, *recall* de 98,91%, *F1-Score* 98,85% e 99,66% de especificidade. Vale ressaltar que a validação externa desempenha um papel fundamental na comparação de arquiteturas CNN, pois simula situações do mundo real, permitindo analisar qual modelo melhor generaliza para novas amostras.

Arquitetura	Acc (%)	Prec (%)	Recall (%)	F1 (%)	Esp (%)
DenseNet201	97,96	97,29	97,92	97,58	99,29
EfficientNetB7	99,01	98,79	98,91	98,85	99,66
ResNet50	97,85	97,23	98,11	97,63	99,23

Tabela 5. Resultado da validação externa da classificação de tumores cerebrais

A Figura 4 exibe a matriz de confusão para o modelo EfficientNetB7 durante a validação externa. Em resumo, o modelo demonstrou um desempenho robusto ao classificar as quatro classes, com um alto número de verdadeiros positivos: 696 instâncias para a classe Meningioma, 1412 instâncias para a classe Glioma, 920 instâncias para a classe Hipofisário e 1491 para a classe sem tumor.

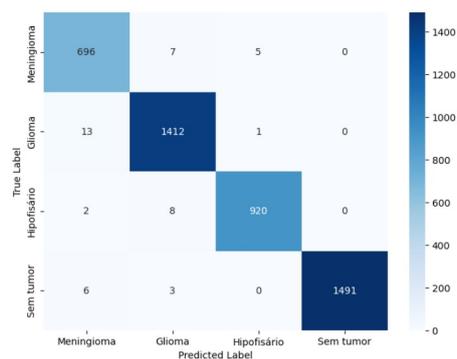


Figura 4. Matriz de Confusão para validação externa da rede EfficientNetB7

Os resultados revelam um potencial promissor nos modelos DenseNet201, EfficientNetB7 e ResNet50 para a classificação de tumores cerebrais. No geral, não foram observadas diferenças estatisticamente significativas entre os três modelos. No entanto, destaca-se que a rede EfficientNetB7 apresentou as melhores taxas para todas as métricas avaliadas. Isso evidencia a capacidade do modelo em classificar de forma precisa os tipos de tumores e diferenciá-los de casos sem tumor.

5. Conclusão e Trabalhos Futuros

Neste trabalho, foi realizada a classificação dos tipos de tumores em Meningioma, Glioma, Hipofisário e casos sem tumor, por meio de imagens de ressonância magnética, baseado em aprendizagem profunda. Para isso, foram realizados diversos experimentos utilizando modelos robustos de Deep Learning.

Para a tarefa de classificação, o estudo apresentou resultados significativos frente ao que é explorado na literatura. Foram realizados testes com cinco redes neurais profundas pré-treinadas: AlexNet, DenseNet201, EfficientNetB7, MobileNetV2 e ResNet50. Nesse contexto, a rede EfficientNetB7 apresentou os melhores resultados para as métricas quantitativas de avaliação, alcançando 97,68% para a acurácia de teste. Logo, os resultados obtidos não apenas validam a abordagem adotada, mas também sugerem possíveis aprimoramentos para otimizações futuras.

Além disso, os resultados do teste de *Nemenyi* confirmam diferenças estatisticamente significativas nas médias das métricas de avaliação entre as arquiteturas AlexNet e MobileNetV2 em comparação com DenseNet201, EfficientNetB7 e ResNet50. Por outro lado, as redes DenseNet201, EfficientNetB7 e ResNet50 apresentaram diferenças estatísticas insignificantes para a métrica acurácia. Portanto, recomenda-se considerar a utilização de uma dessas três redes, devido às taxas consideráveis em termos de métricas de avaliação e validação externa, destacando-se especialmente a EfficientNetB7.

Como trabalhos futuros, deve-se testar outras arquiteturas de redes neurais profundas que sejam leves e eficientes, a fim de desenvolver um sistema de diagnóstico auxiliado por computador (CAD) para detectar, segmentar e classificar tumores cerebrais, por meio de imagens de ressonância magnética com maior precisão e performance. Por fim, o sistema deverá possibilitar a quantificação da área tumoral, a fim de apoiar na determinação do estágio da doença. Logo, o sistema servirá como uma ferramenta de apoio para auxiliar os profissionais de saúde a realizarem diagnósticos precoce com maior precisão.

Referências

- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74.
- Asif, S., Yi, W., Ain, Q. U., Hou, J., Yi, T., and Si, J. (2022). Improving effectiveness of different deep transfer learning-based models for detecting brain tumors from mr images. *IEEE Access*, 10:34716–34730.
- Bhuvaji, S., Kadam, A., Bhumkar, P., Dedge, S., and Kanchan, S. (2020). Brain Tumor Classification (MRI) [Dataset]. Kaggle. <https://www.kaggle.com/dsv/1183165>. Acesso em 01 de março de 2024.
- Bindu, J. H. and Devi, M. U. (2024). Classification of Brain Tumor Images using Segmentation and Transfer Learning. In *2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, pages 225–232.
- Bow, S.-T. (2002). *Pattern Recognition and Image Preprocessing*. Marcel Dekker, Inc., USA, 2nd edition.
- Brown, M. B. and Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American statistical association*, 69(346):364–367.
- Chakrabarty, N. (2017). Brain MRI Images for Brain Tumor Detection [Dataset]. Kaggle. <https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection>. Acesso em 01 de março de 2024.
- Cheng, J. (2017). Brain Tumor Dataset. https://figshare.com/articles/dataset/brain_tumor_dataset/1512427. Acesso em 01 de março de 2024.
- Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., and Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563.
- El-Assiouti, O. S., Hamed, G., El-Saadawy, H., Ebied, H. M., and Khattab, D. (2023). Regioninpaint, cutoff and regionmix: Introducing novel augmentation techniques for enhancing the generalization of brain tumor identification. *IEEE Access*, 11:83232–83250.
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 32(200):675–701.
- 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 *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269.

- Instituto Nacional de Câncer (INCA) (2022). Câncer do sistema nervoso central. <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/sistema-nervoso-central>. Acesso em 01 de março de 2024.
- Islam, M. A., Noshin, S. A., Islam, M. R., Razy, M. F., Antara, S., Reza, M. T., and Parvez, M. Z. (2023). A low parametric cnn based solution to efficiently detect brain tumor cells from ultrasound scans. In *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 1152–1158.
- Kaya, Y. and Gürsoy, E. (2023). A mobilenet-based cnn model with a novel fine-tuning mechanism for covid-19 infection detection. *Soft Computing*, 27(9):5521–5535.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Nemenyi, P. B. (1963). *Distribution-free multiple comparisons*. Princeton University.
- Ottom, M. A., Rahman, H. A., and Dinov, I. D. (2022). Znet: Deep learning approach for 2d mri brain tumor segmentation. *IEEE Journal of Translational Engineering in Health and Medicine*, 10:1–8.
- Padmapriya, S. and Devi, M. G. (2024). Computer-Aided Diagnostic System for Brain Tumor Classification using Explainable AI. In *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, volume 2, pages 1–6.
- Patara, Y., Yadav, S., Pingale, R., and S, K. (2024). Brain Tumor Detection Using Deep Learning(CNN). In *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pages 1–5.
- Raj, J. R. F., Vijayalakshmi, K., Priya, S. K., and Appathurai, A. (2024). Brain tumor segmentation based on kernel fuzzy c-means and penguin search optimization algorithm. *Signal, Image and Video Processing*, 18(2):1793–1802.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611.
- Srivastava, A., Khare, A., and Kushwaha, A. (2023). Brain tumor classification using deep learning framework. In *2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC)*, pages 1–4.
- Tan, M. and Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *International conference on machine learning*, pages 6105–6114. PMLR.
- Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 839–846.
- Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485.