UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Carlos Augusto Gomes Neto

Avaliação de Arquiteturas de Redes Neurais Convolucionais na Identificação de Tumores Cerebrais por Ressonância Magnética

Uberlândia, Brasil

2026

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Carlos Augusto Gomes Neto

Avaliação de Arquiteturas de Redes Neurais Convolucionais na Identificação de Tumores Cerebrais por Ressonância Magnética

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Orientador: Fernanda Maria da Cunha Santos

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2026

Carlos Augusto Gomes Neto

Avaliação de Arquiteturas de Redes Neurais Convolucionais na Identificação de Tumores Cerebrais por Ressonância Magnética

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Trabalho aprovado. Uberlândia, Brasil, 27 de março de 2026:

Fernanda Maria da Cunha Santos
Orientador

Alessandra Aparecida Paulino
Professor

Arthur do Prado Labaki
Professor

Uberlândia, Brasil
2026

Resumo

O câncer cerebral é um tipo agressivo de neoplasia que requer diagnóstico precoce para tratamento eficaz. A ressonância magnética é um exame de diagnóstico por imagem não invasivo, amplamente utilizada na visualização de tumores cerebrais devido à sua capacidade de detalhar a estrutura do cérebro. No entanto, a identificação precoce e precisa dos diferentes tipos de tumores ainda é um desafio devido às similaridades visuais entre eles e à dependência de métodos invasivos. Este trabalho avalia arquiteturas de redes neurais convolucionais para classificar tumores cerebrais em imagens de ressonância magnética, com o objetivo de melhorar a precisão e eficiência do diagnóstico. Foram utilizados os modelos DenseNet201, EfficientNetB0 e ResNet50 com técnicas de Transfer Learning e fine-tuning, aplicados a conjuntos de dados públicos contendo imagens distribuídas em quatro classes: meningioma, glioma, tumor hipofisário e sem tumor. As imagens passaram por etapas, sendo a primeira a de pré-processamento, o que incluiu redimensionamento, conversão para escala de cinza e filtragem bilateral. Durante a etapa de treinamento, foi empregada a técnica de validação cruzada k-fold com o objetivo de avaliar a estabilidade e a capacidade de generalização dos modelos. Após o treinamento, a etapa de avaliação final do modelo proposto foi aplicado a um conjunto de dados público e inédito ao modelo, utilizando as métricas acurácia, precisão, recall, F1-score e AUC para quantificar o desempenho do mesmo. A EfficientNetB0 apresentou acurácia de 75,63%, precisão de 86,52% e F1-score de 73,30%. A ResNet50 obteve acurácia de 75,13% e a maior AUC (92,86%), enquanto a DenseNet201 apresentou acurácia de 73,10% e F1-score de 67,82%. Os modelos apresentaram sensibilidade entre 69,04% e 73,46% e especificidade superior a 91% em todas as arquiteturas. Os resultados demonstram a capacidade dos modelos em classificar tumores cerebrais em imagens de ressonância magnética, sendo a EfficientNetB0, a de melhor desempenho.

Palavras-chave: Tumores Cerebrais, Redes Neurais Convolucionais, Ressonância Magnética.

Lista de ilustrações

Lista de abreviaturas e siglas

BSI	Bacharelado em Sistemas de Informação
CNN	*Convolutional Neural Networks*
DNN	Deep Neural Networks

Sumário

1 Introdução

O câncer cerebral é uma das formas mais agressivas de neoplasia, e uma detecção precoce é fundamental para um tratamento eficaz da doença. Os tumores cerebrais podem ser divididos em algumas categorias, sendo que os mais comuns incluem gliomas, meningiomas e tumores hipofisários. Cada tipo de tumor possui suas próprias características e desafios ao realizar o diagnóstico. A complexidade da estrutura cerebral e a semelhança visual entre os diferentes tipos de tumores tornam a identificação precoce e precisa extremamente desafiadora.

Atualmente, o diagnóstico desses tipos de tumores envolve uma combinação de exames de imagens, testes clínicos e até mesmo biópsias. A ressonância magnética é o método mais utilizado, pois permite visualizar detalhadamente a estrutura do cérebro e também identificar anomalias. A precisão do diagnóstico precoce, ainda enfrenta desafios em identificar os tipos de tumores pela semelhança estrutural entre eles. Com diagnósticos mais rápidos e precisos, pacientes poderiam iniciar tratamentos adequados antes que a doença progredisse, reduzindo custos hospitalares e melhorando a qualidade de vida. Além disso, avanços nessa área impulsionariam a inovação médica e fortaleceriam os sistemas de saúde, gerando impactos positivos.

A identificação precoce e precisa de diferentes tipos de tumores cerebrais continua sendo um grande desafio devido à complexidade estrutural do cérebro e às similaridades visuais entre os tumores. Apesar de avanços em exames de imagem, como a ressonância magnética, o diagnóstico ainda é dependente de métodos demorados e invasivos, como biópsias, e prejudicado com a dificuldade de distinguir certos tipos de tumores. Isso evidencia a necessidade de soluções mais eficazes e inovadoras para melhorar a precisão e a velocidade do diagnóstico, contribuindo para tratamentos mais eficazes e impactos positivos no sistema de saúde.

Nesse contexto, sistemas computacionais têm se destacado como alternativas promissoras, especialmente aqueles baseados em Inteligência Artificial (IA) aplicados à análise de imagens médicas. Técnicas de aprendizado de máquina e, principalmente, de aprendizado profundo (*Deep Learning*) permitem a extração automática de características relevantes a partir das imagens de ressonância magnética, auxiliando na identificação de padrões que podem não ser facilmente perceptíveis ao olho humano. Essas abordagens têm potencial para aumentar a precisão diagnóstica, reduzir o tempo de análise e atuar como ferramentas de apoio à decisão clínica, contribuindo diretamente para a melhoria dos processos na área da saúde.

Diversos estudos recentes vêm explorando o uso de IA na classificação de tumo-

res cerebrais em imagens médicas. O trabalho de Moreira et al. (2024), por exemplo, utiliza arquiteturas de *Transfer Learning*, como DenseNet, EfficientNet e ResNet, alcançando bons resultados na distinção entre diferentes tipos de tumores. Além disso, outras abordagens baseadas em redes neurais convolucionais (CNNs) também têm apresentado desempenho elevado na identificação automática de padrões em imagens de ressonância magnética. Complementando essa linha, a revisão de Cè et al. (2023) destaca que modelos de aprendizado profundo, como CNNs e variações como CapsNet, vão além da classificação, permitindo também análises mais detalhadas, como a delimitação de margens tumorais e a caracterização não invasiva. Esses resultados reforçam o potencial da IA na área e motivam o desenvolvimento de novas abordagens.

1.1 Objetivo Geral

Adaptar e avaliar arquiteturas de redes neurais pré-treinadas e baseadas em aprendizagem profunda para a classificação de diferentes tipos de tumores cerebrais em imagens de ressonância magnética, visando aumentar a precisão e a eficiência no diagnóstico.

1.2 Objetivos Específicos

- Explorar o potencial de arquiteturas de redes neurais convolucionais (CNNs) no diagnóstico de tumores cerebrais.
- Implementar técnicas de pré-processamento de imagens para melhorar a qualidade do conjunto de treinamento.
- Aplicar e ajustar modelos de *Transfer Learning* para acelerar o aprendizado das redes neurais e aumentar a precisão dos resultados.
- Realizar ajustes de hiperparâmetros para garantir a robustez e evitar sobreajuste nos modelos.
- Validar os modelos utilizando bases de dados públicas.

1.3 Justificativa

O trabalho desenvolvido por Moreira et al. (2024) avança na compreensão e no aprimoramento das técnicas de aprendizagem profunda (*deep learning*) e do processamento de dados para propor soluções mais precisas e eficientes na classificação de tumores cerebrais. A metodologia adotada por Moreira et al. (2024) é robusta, pois abrange desde o pré-processamento das imagens, incluindo o aumento do conjunto de dados — que visa

corrigir o desequilíbrio de classes e aumentar a variabilidade dos dados — até a aplicação de técnicas de *Transfer Learning*, que aproveitam modelos que já foram treinados em grandes bases de dados para acelerar o aprendizado das redes neurais responsáveis pela classificação dos tumores e trazer precisão nos resultados.

Assim, com diagnósticos mais rápidos e precisos, pacientes poderiam iniciar tratamentos adequados antes que a doença progredisse, reduzindo custos hospitalares e melhorando a qualidade de vida. Por isso, o trabalho desenvolvido por Moreira et al. (2024) servirá de referência primordial para a progressão deste, acreditando que avanços nessa área impulsionariam a inovação médica e fortaleceriam os sistemas de saúde, gerando impactos positivos.

1.4 Estrutura do Trabalho

Este trabalho está organizado da seguinte forma:

- **Capítulo 2**: são apresentados os fundamentos teóricos relacionados ao problema abordado, incluindo conceitos de aprendizagem profunda e métodos para processamento de imagens médicas. Além disso, serão descritos os trabalhos usados como referência.
- **Capítulo 3**: são descritas as etapas da metodologia utilizada na construção deste estudo, abrangendo o conjunto de dados, as técnicas de pré-processamento, os modelos testados e as métricas de avaliação.
- **Capítulo 4**: são apresentados os resultados obtidos no teste de cada modelo.
- **Capítulo 5**: conclusões e trabalhos futuros são apresentados.

1.5 IA-Generativa

Em conformidade com o Código de Conduta da Sociedade Brasileira de Computação (SBC), declara-se que ferramentas de Inteligência Artificial Generativa (ChatGPT, baseado em modelos da OpenAI) foram utilizadas como apoio na redação, revisão gramatical, aprimoramento do estilo acadêmico e organização textual deste trabalho. Além disso, a ferramenta também foi utilizada como suporte na superação de limitações práticas relacionadas a custo computacional e tempo de treinamento dos modelos, por meio da obtenção de sugestões e orientações. Ressalta-se que a responsabilidade técnica pelo desenvolvimento do software, definição da metodologia, escolha dos algoritmos e interpretação dos resultados permanece integralmente sob responsabilidade dos autores.

2 Fundamentação Teórica

2.1 Pré-processamento de Imagens

O pré-processamento de imagens consiste em transformar os dados brutos de uma imagem original em dados realçados de uma imagem nítida, com o objetivo de permitir que apenas os dados necessários sejam extraídos (BOW, 2002).

Nessa etapa, ocorre o redimensionamento e normalização. Redimensionar uma imagem é o processo de alterar as dimensões físicas (largura e altura) ou a resolução (pixels) de uma imagem digital, com intuito de reamostrar a base de dados (aumento/upscaling ou redução/downscaling). Normalizar é uma técnica que ajusta a faixa de intensidade dos pixels para um intervalo padrão, melhorando o contraste e a consistência visual.

Neste trabalho, as imagens foram convertidas para a escala de cinza, redimensionadas para um tamanho de 224x224 pixels, e também foi aplicada a filtragem bilateral com o objetivo de suavizar as imagens, reduzir ruídos indesejados e preservar a nitidez das bordas (TOMASI; MANDUCHI, 1998).

2.2 Aumento de Dados

O aumento de dados (data augmentation) consiste na aplicação de transformações artificiais em imagens com o objetivo de gerar novas amostras a partir de um conjunto de dados existente, aumentando sua variabilidade e contribuindo para a melhoria do desempenho de modelos de aprendizado de máquina (SHORTEN; KHOSHGOFTAAR, 2019).

Essa técnica é amplamente utilizada em tarefas de visão computacional, especialmente quando o conjunto de dados disponível é limitado. As transformações aplicadas podem incluir rotações, inversões, translações, alterações de escala, entre outras, preservando as características essenciais da imagem enquanto introduzem variações que auxiliam o modelo a generalizar melhor.

Neste trabalho, o aumento de dados foi utilizado durante o treinamento com o objetivo de reduzir o overfitting e melhorar a capacidade de generalização dos modelos, permitindo que as redes neurais fossem expostas a diferentes variações das imagens de ressonância magnética.

2.3 Redes Neurais de Aprendizagem Profunda

As Redes Neurais de Aprendizagem Profunda, também conhecidas como *Deep Neural Networks* (DNNs), são modelos computacionais compostos por múltiplas camadas de neurônios artificiais que são capazes de aprender representações complexas de dados. Esses modelos pertencem ao *Deep Learning*, uma subárea do aprendizado de máquina que se fundamenta no uso de redes neurais com diversas camadas para realizar a extração automática de características relevantes a partir dos dados brutos (ALZUBAIDI et al., 2021).

Diferentemente de abordagens tradicionais de aprendizado de máquina, onde as características precisam ser definidas manualmente, os modelos de aprendizagem profunda possuem a capacidade de aprender essas representações de forma hierárquica. Nesse processo, as primeiras camadas da rede identificam padrões mais simples, enquanto as camadas mais profundas combinam essas informações para representar estruturas mais complexas presentes nos dados.

Existem diferentes tipos de arquiteturas de redes neurais profundas, desenvolvidas para distintos tipos de dados e problemas, como as Redes Neurais Recorrentes (RNNs), utilizadas em dados sequenciais, arquiteturas baseadas em Transformers e modelos generativos, como as GANs. Entre essas arquiteturas, destacam-se as Redes Neurais Convolucionais, amplamente utilizadas em tarefas relacionadas ao processamento e análise de imagens.

2.4 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (*Convolutional Neural Networks* – CNNs) são uma classe de redes neurais profundas projetadas para o processamento de dados com estrutura espacial, como imagens. Essas redes utilizam operações de convolução com o objetivo de extrair características relevantes, como bordas, texturas e padrões visuais presentes nas imagens (ALZUBAIDI et al., 2021).

A convolução é uma operação que consiste na aplicação de pequenos filtros (ou *kernels*) que percorrem a imagem de entrada realizando multiplicações e somas entre os valores dos pixels e os pesos do filtro. Como resultado, são gerados mapas de características (*feature maps*) que destacam padrões específicos presentes na imagem. Durante o processo de treinamento, os valores desses filtros são ajustados automaticamente pelo modelo, permitindo que a rede aprenda quais são os padrões mais relevantes para a classificação.

A arquitetura de uma CNN é composta por diferentes tipos de camadas. Destacam-se as camadas convolucionais, responsáveis pela extração de características locais por meio da aplicação de filtros sobre as imagens. Após as convoluções, é comum a utilização de

funções de ativação não lineares, como a ReLU (*Rectified Linear Unit*), que introduzem não linearidade ao modelo e permitem que a rede aprenda representações mais complexas dos dados. Além disso, são utilizadas camadas de *pooling*, que realizam a redução da dimensionalidade das representações geradas, mantendo as informações mais relevantes. Essa redução contribui para diminuir o custo computacional e reduzir o risco de *overfitting*. Ao final da rede, camadas totalmente conectadas (*fully connected*) são utilizadas para realizar a etapa de classificação com base nas características extraídas ao longo do processamento.

Outro aspecto importante das CNNs é o aprendizado hierárquico de características. Nas camadas iniciais da rede são identificados os padrões mais simples, como contrastes e bordas. À medida que a informação percorre camadas mais profundas, a rede passa a identificar padrões mais complexos e abstratos, como formas e estruturas presentes nas imagens.

De forma geral, as CNNs podem ser compreendidas como um tipo específico de rede neural profunda, pertencente ao campo do *Deep Learning*, destacando-se pela elevada eficácia em aplicações de visão computacional, como reconhecimento de objetos, detecção de padrões e classificação de imagens médicas.

A Figura 4 mostra um exemplo da arquitetura de uma CNN utilizada para a classificação de imagens.

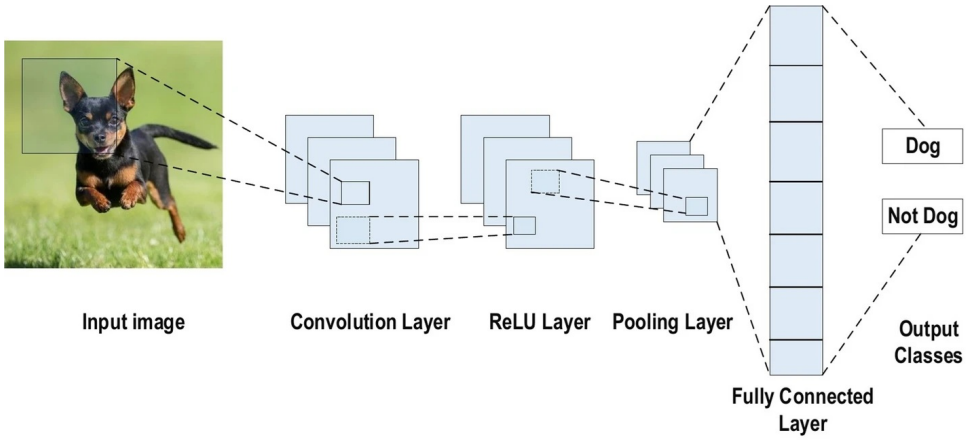


Figura 1 – Exemplo de arquitetura de uma Rede Neural Convolucional. Fonte: (ALZUBAIDI et al., 2021).

Diante da elevada complexidade dessas arquiteturas e da grande quantidade de dados necessária para treiná-las do zero, técnicas como o *Transfer Learning* têm sido amplamente utilizadas para aproveitar o conhecimento previamente aprendido por modelos treinados em grandes bases de dados.

2.5 *Transfer Learning*

Após o processo de pré-processamento de imagens e aumento de dados, um modelo de classificação é responsável por definir a probabilidade de uma entrada pertencer a uma determinada classe. Contudo, treinar uma arquitetura de Rede Neural Convolucional (CNN) do zero demanda muito tempo e recursos computacionais. Para atenuar essa dificuldade, a técnica de *Transfer Learning* é comumente aplicada, utilizando modelos CNN pré-treinados em grandes conjuntos de dados e aproveitando o conhecimento já adquirido (ALZUBAIDI et al., 2021), (KAYA; GURSOY, 2023).

Nesse contexto, as arquiteturas DenseNet201 (HUANG et al., 2017), EfficientNetB0 (TAN; LE, 2019) e ResNet50 (HE et al., 2016), todas com pesos pré-treinados no conjunto de dados ImageNet, foram empregadas para a classificação dos tipos de tumores. Essa abordagem acelera o aprendizado do modelo e melhora sua precisão.

2.6 Validação Cruzada

A Validação Cruzada é uma técnica essencial para avaliar a capacidade de generalização do modelo, garantindo que ele mantenha bom desempenho ao lidar com diferentes dados. Um modelo que performa bem apenas nos dados de treino pode sofrer de *overfitting*, o que compromete sua eficácia em cenários reais.

A abordagem *k-fold cross-validation* foi utilizada, na qual o conjunto de dados é dividido em k subconjuntos (*folds*). A cada iteração, um subconjunto é utilizado como teste, enquanto os $k-1$ restantes são usados para treino. Esse processo se repete k vezes, e o desempenho final do modelo é calculado como a média dos desempenhos de cada *fold*.

Essa metodologia oferece uma avaliação robusta, especialmente útil em conjuntos de dados limitados, permitindo identificar padrões gerais e reduzir o *overfitting*. Dessa forma, garante-se que o modelo seja mais preciso e generalizado quando exposto a novos e diferentes dados.

2.7 Trabalhos Relacionados

Os estudos sobre a classificação de tumores cerebrais por meio de imagens de ressonância magnética apresentam diversas abordagens para aprimorar a precisão do diagnóstico. Por exemplo, o trabalho de Moreira et al. (2024) sugere um fluxo de trabalho que utiliza arquiteturas pré-treinadas de CNNs, combinando técnicas robustas de pré-processamento, aumento de dados e validação cruzada k-fold para garantir maior precisão e confiabilidade na classificação. Já Yapici Rukiye Karakis (2022) adota uma estratégia inovadora de geração de dados sintéticos com CycleGAN para contornar limitações

relacionadas ao tamanho do conjunto de dados.

No estudo de Moreira et al. (2024), são exploradas arquiteturas de *Transfer Learning* como DenseNet201, EfficientNetB7 e ResNet50, aplicando ajuste fino de hiperparâmetros via Grid Search, o que otimiza o desempenho dos modelos. A base de dados principal compreendeu 3.264 imagens classificadas em quatro categorias: Meningioma, Glioma, Hipofisário e casos sem tumor. No pré-processamento, essa abordagem incluiu conversão das imagnes para escala de cinza, redimensionamento, filtragem bilateral e equalização adaptativa de histograma (CLAHE), visando melhorar a qualidade e o contraste das imagens. Além disso, os modelos foram avaliados por meio de validação externa com conjuntos de dados independentes, o que contribuiu para demonstrar sua capacidade de generalização em cenários distintos. Os resultados obtidos indicaram alto desempenho das arquiteturas empregadas, com métricas elevadas de acurácia, precisão e recall, evidenciando a eficácia da abordagem proposta para a classificação de imagens de ressonância magnética cerebral. Destaca-se que a arquitetura EfficientNetB7 obteve os melhores resultados, alcançando 99,01% de acurácia na validação externa, além de apresentar valores elevados nas demais métricas avaliativas, reforçando seu potencial para aplicação em diagnósticos automatizados de tumores cerebrais.

Por sua vez, Yapici Rukiye Karakis (2022) concentra-se na criação de imagens sintéticas com o objetivo de melhorar a classificação em conjuntos de dados de tamanhos limitados. Para isso, utilizam a arquitetura Cycle Consistent Generative Adversarial Network (CycleGAN) para gerar imagens realistas de ressonância magnética com três tipos de tumores cerebrais (glioma, meningioma e pituitário), respeitando as características anatômicas específicas de cada corte axial, sagital e coronal. As imagens sintéticas foram combinadas com os dados originais para treinar um modelo de classificação baseado na arquitetura DenseNet121. Essa estratégia superou as limitações da escassez de dados, aumentando a diversidade do conjunto de treinamento e prevenindo o overfitting. Como resultado, o modelo treinado com a combinação de dados reais e sintéticos obteve melhorias significativas na acurácia, com valores médios que aumentaram de aproximadamente 96% para mais de 99% nas três classes tumorais, além de apresentar melhor desempenho em métricas como precisão, sensibilidade, especificidade e área sob a curva (AUC).

Ambos os trabalhos têm como objetivo aumentar a acurácia e diminuir a má interpretação entre os tipos de tumores, mas diferem nas metodologias: o primeiro foca em técnicas avançadas de treinamento, pré-processamento e validação rigorosa, enquanto o segundo investiga novas formas de aumentar a variabilidade dos dados por meio de geração sintética.

A análise de alguns estudos revela algumas lacunas significativas que este trabalho busca abordar:

- falta de padronização na metodologia: os estudos desenvolvidos por Moreira et al. (2024) e Yapici Rukiye Karakis (2022) apresentaram bons resultados, porém utilizam diferentes métodos na construção do classificador, conjuntos de dados e métricas, dificultando uma comparação direta entre os modelos pré-treinados.

- uso limitado de métricas de avaliação: A maioria dos estudos prioriza métricas tradicionais, como acurácia, sem explorar de forma aprofundada métricas mais robustas, como a AUC, que avaliam a capacidade de separação entre classes.

- dependência de conjuntos de dados limitados: Apesar do uso de técnicas como geração de dados sintéticos, ainda há desafios relacionados à variabilidade e generalização dos modelos.

O trabalho proposto busca aperfeiçoar os estudos já realizados ao unir técnicas avançadas de *Transfer Learning* com métodos de geração de dados sintéticos, visando superar as limitações identificadas. Além disso, será realizada uma análise comparativa mais abrangente entre diferentes arquiteturas de redes neurais, considerando múltiplas métricas de avaliação.

Destaca-se, nesse contexto, a utilização da métrica AUC (*Area Under the Curve*), que complementa métricas tradicionais e permite uma avaliação mais robusta da capacidade dos modelos em distinguir entre as diferentes classes.

A combinação das abordagens discutidas em Moreira et al. (2024) e Yapici Rukiye Karakis (2022), com adaptações específicas com o intuito de investigar de forma mais aprofundada a aplicação de técnicas de aprendizado profundo na classificação de tumores cerebrais em imagens de ressonância magnética.

3 Metodologia Proposta

Neste capítulo serão apresentadas todas as etapas que compuseram o desenvolvimento do modelo computacional para classificação de tumores cerebrais em imagens de ressonância magnética. A Figura 2 apresenta o fluxograma da metodologia proposta, descrevendo as principais etapas do processo, desde a preparação dos dados até a avaliação do desempenho dos modelos.

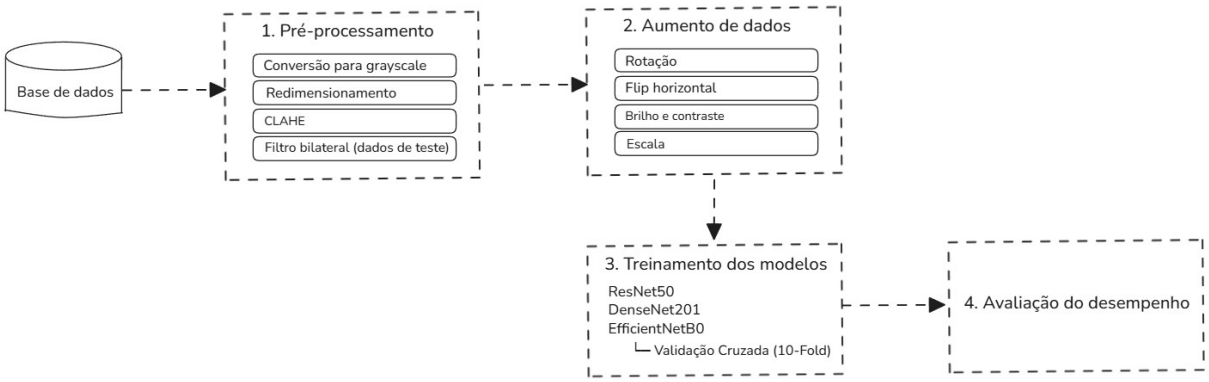


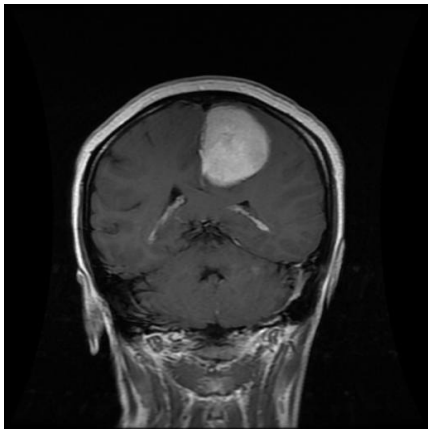
Figura 2 – Fluxograma das etapas do modelo computacional

3.1 Base de Dados

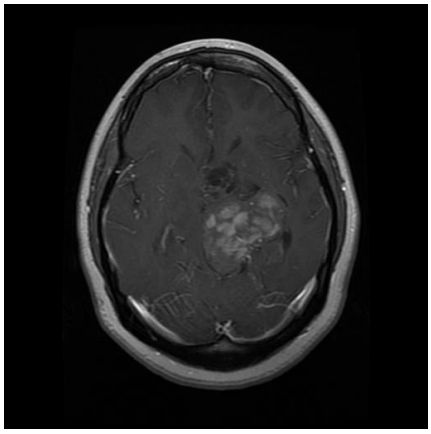
A etapa inicial da metodologia consiste na preparação do conjunto de dados utilizado para o treinamento e, posteriormente, para o teste dos modelos de classificação. Foram utilizadas imagens de ressonância magnética (MRI) de tumores cerebrais provenientes de bases de dados públicas amplamente utilizadas na literatura científica.

O principal conjunto de dados utilizado no treinamento dos modelos de classificação foi o dataset disponibilizado por Bhuvaji et al. (2020). Esse conjunto contém 3264 imagens de ressonância magnética com contraste aprimorado, classificadas em quatro categorias distintas: meningioma (937 imagens), glioma (926 imagens), tumor hipofisário (901 imagens) e imagens sem tumor (500 imagens).

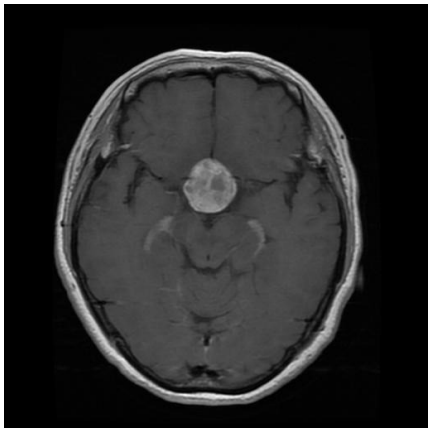
As imagens da Figura 3 são exemplos de imgens obtidas pelos exames de ressonância magnética, correspondentes às quatro classes consideradas neste trabalho: (1) meningioma, (2) glioma, (3) tumor hipofisário e (4) sem tumor.



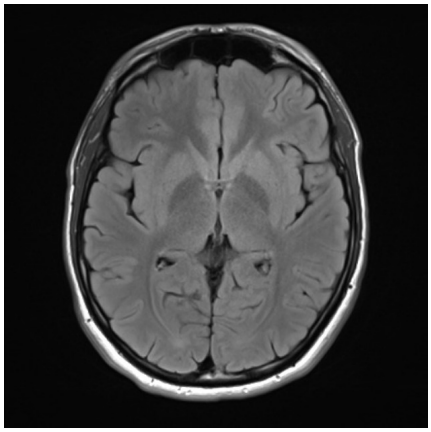
(a) Meningioma



(b) Glioma



(c) Tumor hipofisário



(d) Sem tumor

Figura 3 – Exemplos de imagens das classes presentes no dataset utilizado no treinamento.

Além do dataset utilizado no treinamento, foram utilizados outros para a realização da fase de teste, com o objetivo de avaliar a capacidade de generalização dos modelos quando expostos a imagens provenientes de outras fontes. Um dos conjuntos utilizados foi o Figshare Brain Tumor Dataset (CHENG, 2017), que contém 3064 imagens de ressonância magnética obtidas de 233 pacientes. Essas imagens estão classificadas em três categorias: meningioma (708 imagens), glioma (1426 imagens) e tumor hipofisário (930 imagens).

Além das imagens de Cheng (2017), também foi utilizado na fase de teste o dataset Br35H (CHAKRABARTY, 2017), originalmente composto por 3000 imagens de ressonância magnética, sendo 1500 classificadas como com tumor e 1500 como sem tumor. No presente trabalho, foram utilizadas apenas as 1500 imagens sem tumor, com o objetivo de complementar o conjunto de teste.

A Tabela 1 apresenta um resumo das características dos datasets utilizados neste estudo, incluindo o número de classes, a quantidade de imagens disponíveis e a finalidade de uso de cada conjunto de dados no experimento.

Tabela 1 – Datasets utilizados no estudo

Dataset	**Classes**	**Nº de imagens**	**Uso**	**Quantidade por classe**
Bhuvaji et al.	4	3264	Treinamento	937 (Meningioma) 926 (Glioma) 901 (Tumor hipofisário) 500 (sem tumor)
Figshare	3	3064	Teste	708 (Meningioma) 1426 (Glioma) 930 (Tumor hipofisário)
Br35H	2	1500	Teste	1500 (sem tumor)

3.2 Pré-processamento dos Dados

A etapa de pré-processamento tem o objetivo de padronizar as imagens e melhorar a qualidade para a classificação dos modelos de aprendizado profundo. Inicialmente, todas as imagens foram convertidas para escala de cinza, reduzindo a dimensionalidade dos dados e mantendo as informações estruturais mais relevantes presentes nas imagens médicas.

Em seguida, as imagens foram redimensionadas para o tamanho fixo de 224×224 pixels, garantindo compatibilidade com as arquiteturas de CNNs utilizadas no treinamento. A padronização do tamanho das imagens também contribui para tornar o processo de treinamento mais estável.

Para melhorar o contraste das imagens, foi aplicada condicionalmente a técnica CLAHE (*Contrast Limited Adaptive Histogram Equalization*) em imagens com baixa variabilidade de intensidade. Esse método permite aumentar o contraste local da imagem sem amplificar excessivamente o ruído, sendo amplamente utilizado em etapas de pré-processamento para melhorar a qualidade das imagens utilizadas em modelos de aprendizado profundo (ALZUBAIDI et al., 2021). A aplicação do CLAHE foi condicionada ao desvio padrão da imagem, permitindo sua utilização apenas em casos de baixa variabilidade de intensidade.

No conjunto de teste, também foi aplicado um filtro bilateral com o objetivo de reduzir ruídos, preservando as bordas e estruturas anatômicas relevantes. Esse filtro realiza uma suavização considerando não apenas a proximidade entre os pixels, mas também a similaridade de intensidade, de modo que apenas pixels próximos e com valores semelhantes influenciem no resultado. Assim, o ruído é reduzido sem borrar as bordas, mantendo os contornos importantes da imagem.

Por fim, após a execução de todas essas funções, todas as imagens foram normalizadas no intervalo $[0, 1]$, preparando os dados para o treinamento e teste dos modelos de classificação.

3.3 Técnica para Aumento dos Dados

Em aplicações de aprendizado profundo voltadas para imagens médicas, a quantidade limitada de dados pode comprometer a capacidade de generalização dos modelos. Nesse contexto, técnicas de aumento de dados (*data augmentation*) são utilizadas com o objetivo de ampliar artificialmente o *dataset*.

O aumento de dados consiste na geração de novos dados a partir das imagens originais por meio de transformações controladas. Essas novas imagens preservam as características estruturais relevantes presentes nos dados originais. Essa etapa foi aplicada exclusivamente ao conjunto de treinamento com o objetivo de aumentar a diversidade das amostras utilizadas durante o processo de aprendizado e reduzir o risco de *overfitting*.

As transformações aplicadas incluem espelhamento horizontal, rotações aleatórias no intervalo de $-30°$ a $30°$, variações aleatórias de brilho e contraste e transformações de escala. Essas operações permitem gerar diferentes representações das mesmas imagens, simulando variações que podem ocorrer durante a realização de exames de ressonância magnética, como pequenas alterações no posicionamento do paciente ou diferenças nas condições de captura das imagens.

A Figura 4 apresenta exemplos dessas transformações aplicadas às imagens do conjunto de treinamento, ilustrando como uma mesma imagem original pode gerar múltiplas variações, contribuindo para o aumento da diversidade dos dados utilizados no treinamento do modelo.

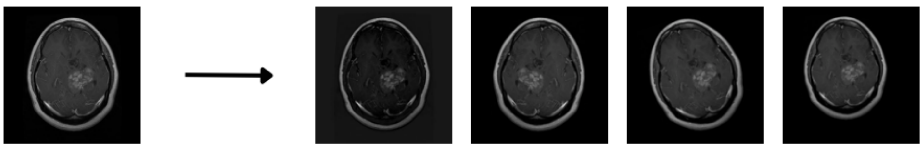


Figura 4 – Transformações aplicadas nos dados de treinamento.

O conjunto de teste não recebeu imagens geradas pela técnicas de aumento de dados, garantindo que a validação dos modelos seja realizada apenas por dados inéditos e não repetidos.

A Tabela 2 apresenta a quantidade de imagens antes e após a aplicação da técnicade aumento de dados.

Tabela 2 – Número de imagens antes e após o aumento de dados

Etapa	**Imagens**	**Descrição**
Dataset original	3264	Imagens originais
Após pré-processamento e aumento	16320	1 imagem + 4 aumentos

3.4 Configuração e Treinamento dos Modelos Pré-treinados

3.4.1 Arquiteturas Utilizadas e Transfer Learning

Foram utilizadas arquiteturas de CNNs profundas amplamente empregadas em tarefas de classificação de imagens, são elas: ResNet50 (HE et al., 2016), DenseNet201 (HUANG et al., 2017) e EfficientNetB0 (TAN; LE, 2019). Esses modelos foram utilizados como extratores de características, já com pesos pré-treinados, oriundos do *dataset* ImageNet (DENG et al., 2009).

Cada uma dessas arquiteturas possui uma forma diferente de organizar e reaproveitar as informações ao longo da rede. A ResNet50, por exemplo, utiliza conexões residuais (*skip connections*), que permitem contornar determinadas camadas, ajudando a manter o fluxo de informação e facilitando o treinamento de redes mais profundas. Já a DenseNet201 vai além nesse reaproveitamento, conectando cada camada com todas as anteriores, o que favorece o compartilhamento de características e reduz redundâncias. Por outro lado, a EfficientNetB0 segue uma proposta diferente, focando em eficiência: ela utiliza um escalonamento balanceado da rede (*compound scaling*), ajustando profundidade, largura e resolução de forma conjunta para obter bom desempenho com menor custo computacional. Essas três arquiteturas foram escolhidas seguindo os princípios do trabalho desenvolvido por Moreira et al. (2024).

O uso de modelos pré-treinados possibilita a exploração de características visuais já aprendidas, diminuindo a demanda por grandes quantidades de dados para treinamento e agilizando o processo de convergência do modelo.

Neste estudo, os modelos pré-treinados não foram utilizados exatamente em sua configuração original. As camadas convolucionais das arquiteturas foram empregadas como extratores de características, aproveitando os padrões visuais aprendidos durante o treinamento no ImageNet, como bordas, texturas e formas presentes nas imagens. Além disso, a entrada das redes foi ajustada para aceitar imagens com apenas um canal, pois as imagens do conjunto de dados foram processadas em escala de cinza. Nos modelos originais treinados no ImageNet, a entrada é composta por imagens RGB com três canais de cor. Dessa forma, foi necessário adaptar a etapa de entrada para permitir o processamento de imagens em escala de cinza, mantendo a compatibilidade com a arquitetura das redes.

Além disso, a última camada neural das arquiteturas dos modelos pré-treinados, responsável pela classificação, foi substituída por uma nova camada densa com função de ativação *softmax*, configurada para realizar a classificação em quatro categorias de tumores cerebrais.

Também foi aplicado o processo de *fine-tuning*. Nesse procedimento, parte das camadas superiores das redes neurais é liberada para treinamento, permitindo que os

pesos originalmente aprendidos no ImageNet sejam ajustados ao novo problema. Dessa forma, o modelo consegue adaptar o seu conhecimento prévio às novas características e reconhecer os padrões específicos presentes nas imagens de tumores cerebrais.

3.4.2 Hiperparâmetros de Treinamento

Durante o treinamento dos modelos foram definidos alguns hiperparâmetros importantes, como a taxa de aprendizagem, o otimizador e o tamanho do lote (*batch size*). Esses parâmetros influenciam diretamente o processo de treinamento e o desempenho final das redes neurais no processo de classificação.

A definição desses valores foi realizada com base em experimentos empíricos conduzidos ao longo do desenvolvimento do trabalho (GOODFELLOW; BENGIO; COURVILLE, 2016). Como diferentes arquiteturas podem apresentar comportamentos distintos durante o treinamento, foi necessário realizar pequenos ajustes para cada modelo avaliado.

A taxa de aprendizagem foi definida com valores variando entre 5×10^{-5} e 1×10^{-4}. Essa escolha é comum em cenários de *transfer learning*, pois permite que os modelos pré-treinados ajustem seus pesos gradualmente ao novo conjunto de dados sem alterar de forma abrupta as representações previamente aprendidas no dataset ImageNet.

O tamanho do lote (*batch size*) também variou entre os modelos. Para as arquiteturas ResNet50 e DenseNet201 foi utilizado um *batch size* de 8 imagens por iteração. Esse valor foi adotado principalmente devido a limitações de memória da GPU utilizada durante os experimentos, além de ter apresentado comportamento estável durante o treinamento.

Já para a EfficientNetB0 foi possível utilizar um *batch size* maior, de 32 imagens. Isso ocorreu porque essa arquitetura possui uma estrutura mais eficiente em termos de utilização de memória, permitindo o processamento de um número maior de imagens por iteração.

Em relação aos otimizadores, diferentes estratégias foram utilizadas para cada modelo com base nos resultados observados durante os testes preliminares. Inicialmente, foram avaliados diferentes otimizadores para cada arquitetura, considerando critérios como estabilidade do treinamento, velocidade de convergência e desempenho nas métricas de validação.

O otimizador Adam foi utilizado na EfficientNetB0 por apresentar melhor desempenho geral durante os testes, além de ser amplamente empregado em tarefas de visão computacional devido à sua capacidade adaptativa de ajuste das taxas de aprendizado. Na DenseNet201 foi utilizado o AdamW, que incorpora um mecanismo de regularização por desacoplamento do *weight decay*, contribuindo para a redução do sobreajuste observado durante os experimentos. Já para a ResNet50 foi adotado o RMSprop, que demonstrou

maior estabilidade no processo de treinamento dessa arquitetura específica, apresentando melhores resultados em comparação aos demais otimizadores avaliados.

Dessa forma, a escolha dos otimizadores foi guiada por uma abordagem empírica, baseada na análise comparativa dos resultados obtidos durante a fase experimental. A Tabela 3 apresenta, resumidamente, os hiperparâmetros utilizados no treinamento de cada modelo avaliado neste trabalho.

Tabela 3 – Hiperparâmetros utilizados no treinamento dos modelos

Modelo	**Taxa de Aprendizagem**	**Otimizador**	**Batch Size**
ResNet50	0,00005	RMSprop	8
DenseNet201	0,00005	AdamW	8
EfficientNetB0	0,0001	Adam	32

3.4.3 Estratégias de Regularização e Controle do Treinamento

Durante o treinamento dos modelos foram adotadas estratégias voltadas à melhoria da capacidade de generalização das redes neurais e à redução do risco de *overfitting*. Entre as principais técnicas utilizadas destacam-se o uso de *dropout*, a parada antecipada do treinamento (*early stopping*) e o ajuste adaptativo da taxa de aprendizagem.

Camadas de *dropout* foram inseridas nas camadas densas finais das redes neurais com o objetivo de reduzir a coadaptação entre neurônios durante o treinamento. Esse mecanismo desativa aleatoriamente uma fração das unidades da rede a cada iteração, contribuindo para que o modelo aprenda representações mais robustas.

Também foi utilizado o mecanismo de parada antecipada (*early stopping*), que monitora o desempenho do modelo nas métricas de validação e interrompe o treinamento caso não haja melhoria após um determinado número de épocas. Dessa forma, evita-se o treinamento excessivo e reduz-se o risco de sobreajuste.

Além disso, foi empregado o método *ReduceLROnPlateau*, responsável por reduzir automaticamente a taxa de aprendizagem quando a métrica monitorada deixa de apresentar melhoria. Essa estratégia auxilia na estabilização do processo de otimização e permite um refinamento mais gradual dos pesos do modelo.

Para lidar com possíveis desequilíbrios entre as classes do conjunto de dados, também foram utilizados pesos de classe (*class weights*) durante o treinamento, aumentando a contribuição das classes menos representadas no cálculo da função de perda.

Adicionalmente, foi aplicada a técnica de *fine-tuning* nas arquiteturas pré-treinadas. As redes convolucionais foram inicializadas com pesos previamente treinados, mantendo parte de suas camadas congeladas enquanto as camadas superiores foram liberadas para treinamento.

Essa abordagem permite preservar representações visuais mais gerais aprendidas em grandes bases de dados, ao mesmo tempo em que ajusta as camadas mais profundas às características específicas do conjunto de imagens utilizado neste trabalho. Durante esse processo, camadas de normalização em lote (*Batch Normalization*) permaneceram congeladas, seguindo boas práticas em cenários de *transfer learning*, já que a atualização dessas camadas pode introduzir instabilidades quando o conjunto de dados é limitado.

3.5 Validação Cruzada

Para avaliar o desempenho dos modelos de classificação e reduzir o risco de sobreajuste (*overfitting*), foi utilizada a técnica de validação cruzada do tipo *K-Fold*. Nesse método, o conjunto de dados de treinamento é dividido em k subconjuntos de tamanho aproximadamente igual, denominados *folds*.

Neste trabalho foi adotado $k = 10$, o que significa que o conjunto de treinamento foi dividido em dez partes. Em cada iteração do processo, aproximadamente 90% dos dados (nove *folds*) são utilizados para o treinamento do modelo, enquanto os 10% restantes (um *fold*) são utilizados para validação. Esse processo é repetido dez vezes, de forma que cada subconjunto seja utilizado exatamente uma vez como conjunto de validação.

Ao final das dez iterações, é calculada a média das métricas de desempenho obtidas em cada *fold*, fornecendo uma estimativa mais robusta do desempenho do modelo.

A escolha de $k = 10$ baseia-se em sua ampla utilização na literatura de aprendizado de máquina, sendo considerado um bom equilíbrio entre custo computacional e confiabilidade da estimativa de desempenho (KOHAVI, 1995; HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Valores menores de k podem gerar estimativas mais instáveis, enquanto valores muito elevados aumentam significativamente o tempo de treinamento, especialmente em modelos de aprendizado profundo.

Assim, a utilização do *10-Fold Cross-Validation* permite que o modelo seja treinado e validado em diferentes partições dos dados, contribuindo para uma avaliação mais consistente da capacidade de generalização do modelo.

3.6 Avaliação de Desempenho

A avaliação do desempenho dos modelos de classificação foi realizada por meio de um conjunto de métricas amplamente utilizadas em tarefas de classificação de imagens médicas. Essas métricas permitem analisar diferentes aspectos do desempenho do modelo, especialmente em cenários com múltiplas classes.

Durante a etapa de teste, as predições geradas pelo modelo foram comparadas

com os rótulos reais das imagens pertencentes ao conjunto de teste. A partir dessas informações foi construída a matriz de confusão, a qual permite identificar corretamente os verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos para cada classe. Com base nesses valores, as seguintes métricas são calculadas, com o intuito de avaliar o desempenho das arquiteturas neurais:

- **Acurácia (Accuracy)**: representa a proporção total de classificações corretas realizadas pelo modelo em relação ao número total de amostras avaliadas:

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

- **Precisão (Precision)**: mede a proporção de amostras classificadas como pertencentes a uma determinada classe que realmente pertencem a essa classe:

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

- **Revocação (Recall)**: indica a capacidade do modelo de identificar corretamente as amostras pertencentes a uma determinada classe:

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

- **F1-Score**: corresponde à média harmônica entre precisão e revocação:

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

- **AUC (Area Under the ROC Curve)**: avalia a capacidade do modelo em distinguir corretamente as diferentes classes, considerando a relação entre a taxa de verdadeiros positivos (TPR) e a taxa de falsos positivos (FPR):

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN}$$

- **Especificidade (Specificity)**: mede a capacidade do modelo de identificar corretamente as amostras negativas:

$$Specificity = \frac{TN}{TN + FP}$$

As métricas de precisão, revocação e F1-score foram calculadas utilizando a média do tipo *macro*, garantindo que todas as classes contribuam igualmente para o resultado final, independentemente do número de amostras em cada classe.

Durante a etapa de validação cruzada com $k = 10$, essas métricas foram calculadas individualmente para cada *fold*. Ao final do processo, os resultados foram agregados por

meio da média entre os dez *folds*, fornecendo uma estimativa mais robusta e estável do desempenho dos modelos.

Com base nos resultados obtidos na validação cruzada, foi selecionada a melhor configuração de modelo e hiperparâmetros. Em seguida, realizou-se o treinamento final utilizando todo o conjunto de dados de Bhuvaji et al. (2020).

Posteriormente, o modelo final foi avaliado em um conjunto de teste independente, composto por imagens provenientes dos datasets Figshare e Br35H. Essa etapa teve como objetivo analisar a capacidade de generalização do modelo quando aplicado a dados provenientes de fontes distintas daquelas utilizadas durante o treinamento.

Os resultados obtidos no conjunto de teste foram apresentados por meio das mesmas métricas utilizadas na validação, permitindo uma comparação direta entre o desempenho durante o treinamento e a performance em dados não vistos pelo modelo.

4 Resultados

4.1 Resultados do Modelo Proposto

As arquiteturas neurais ResNet50, DenseNet201 e EfficientNetB0 foram implementadas pelo modelo proposto para a tarefa de classificação de imagens cerebrais, obtidas pelo exame ressonância magnética. Diferentemente do trabalho de referência (Moreira et al. (2024)), que utilizou a EfficientNetB7, optou-se pela EfficientNetB0, versão mais compacta da família EfficientNet, visando reduzir a complexidade computacional e manter viabilidade de treinamento com os recursos disponíveis.

Durante a etapa de treinamento, os modelos foram submetidos a um processo de validação cruzada, com o objetivo de estimar de forma mais robusta o desempenho das arquiteturas durante o treinamento, reduzindo a influência de variações na divisão dos dados. Os resultados desse processo gerou os valores das matrizes de confusões, com os quais calculou-se as métricas acurácia, precisão, recall, F1-score, especificidade e AUC para cada arquitetura neural. A Tabela 4 mostra os valores das métricas que descrevem o desempenho das redes neurais após o treinamento.

Tabela 4 – Resultados gerados pelo modelo proposto na etapa de treinamento.

Arquitetura	**Acc (%)**	**Prec (%)**	**Recall (%)**	**F1 (%)**	**Esp (%)**	**AUC (%)**
DenseNet201	74,18	82,68	70,23	67,98	92,33	93,45
EfficientNetB0	75,93	87,11	73,49	73,32	92,93	93,72
ResNet50	76,21	85,44	72,22	71,92	92,11	92,91

Depois a etapa de treinamento, as redes neurais ResNet50, DenseNet201 e EfficientNetB0 foram testadas, utilizando as bases de dados Figshare Brain Tumor Dataset (CHENG, 2017) e Br35H (CHAKRABARTY, 2017). A Tabela 5 apresenta os valores das métricas que indicam o desempenho de cada arquitetura neural na etapa de teste.

A EfficientNetB0 apresentou a melhor acurácia geral (75,63%), além do maior F1-score (0,73) e da maior precisão (86,52%), demonstrando melhor equilíbrio entre a detecção correta de tumores e o controle de falsos positivos. A ResNet50 apresentou desempenho semelhante (acurácia de 75,13%), com destaque para a maior área sob a curva ROC (AUC = 0,9286), indicando elevada capacidade discriminativa global. Já a DenseNet201 apresentou desempenho inferior em comparação às demais arquiteturas, com acurácia de 73,10% e F1-score de 0,68.

Tabela 5 – Resultados gerados pelo modelo proposto no conjunto de teste.

Arquitetura	**Acc (%)**	**Prec (%)**	**Recall (%)**	**F1 (%)**	**Esp (%)**	**AUC (%)**
DenseNet201	73,10	80,68	69,04	67,82	91,03	91,96
EfficientNetB0	75,63	86,52	73,46	73,30	91,88	92,52
ResNet50	75,13	83,63	72,08	71,79	91,71	92,86

Em termos de sensibilidade (recall), os modelos variaram entre 69 e 73, evidenciando capacidade moderada de detecção de casos positivos. Por outro lado, a especificidade manteve-se elevada em todas as arquiteturas (acima de 91), indicando bom desempenho na identificação correta de exames sem tumor.

De forma geral, os resultados demonstram que, embora os modelos apresentem boa capacidade de separação entre as classes (AUC superior a 0,91 em todas as arquiteturas), ainda há espaço para aprimoramento na sensibilidade, especialmente considerando a relevância clínica da detecção de tumores cerebrais.

4.2 Análise Comparativa

O estudo de Moreira et al. (2024) avaliou o desempenho das arquiteturas DenseNet201, EfficientNetB7 e ResNet50 por meio de uma validação externa utilizando conjuntos de dados não empregados no treinamento (Figshare e Br35H). Esses bancos de dados públicos contêm imagens de ressonância magnética classificadas em diferentes categorias de tumores cerebrais e casos sem tumor.

Nesse procedimento, os modelos previamente treinados foram aplicados nos novos conjuntos de dados para avaliar a capacidade de generalização das redes neurais em amostras independentes, simulando cenários mais próximos de aplicações reais.

As métricas utilizadas para avaliação incluíram acurácia, precisão, recall, F1-score e especificidade. Além disso, o estudo utilizou matrizes de confusão para analisar detalhadamente o desempenho dos modelos em cada classe.

Os resultados reportados por Moreira et al. (2024) apresentaram valores bastante elevados em todas as métricas analisadas, como pode ser lido na Tabela 6. A arquitetura EfficientNetB7 obteve o melhor desempenho geral, alcançando acurácia de 99,01%, seguida pela DenseNet201 (97,96%) e pela ResNet50 (97,85%). Apesar dessas diferenças numéricas, o estudo aponta que não foram observadas diferenças estatisticamente significativas entre os modelos avaliados.

Tabela 6 – Comparação entre os resultados obtidos durante o teste neste trabalho e no estudo de referência.

Arquitetura	**Acc (%)**	**Prec (%)**	**Recall (%)**	**F1 (%)**	**Esp (%)**
Resultados obtidos neste trabalho					
DenseNet201	73,10	80,68	69,04	67,82	91,03
EfficientNetB0	75,63	86,52	73,46	73,30	91,88
ResNet50	75,13	83,63	72,08	71,79	91,71
Resultados reportados no estudo de referência					
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

Em comparação com os resultados obtidos neste trabalho, observa-se que as métricas alcançadas foram consideravelmente inferiores às reportadas no estudo de referência. Enquanto o trabalho de Moreira et al. (2024) apresentou acurácias superiores a 97%, os modelos avaliados neste estudo obtiveram acurácias entre aproximadamente 73% e 75%.

Essa diferença pode estar relacionada a diversos fatores. Entre eles destacam-se possíveis diferenças nas estratégias de treinamento e validação adotadas, bem como nas limitações de recursos computacionais disponíveis para o treinamento dos modelos. Além disso, neste trabalho foi utilizada a arquitetura EfficientNetB0, uma versão mais compacta da família EfficientNet, enquanto o estudo de referência empregou a EfficientNetB7, que possui maior número de parâmetros e maior capacidade representacional.

Outro fator relevante refere-se à estratégia de ajuste de hiperparâmetros adotada. No estudo de Moreira et al. (2024), foi utilizada a técnica de *Grid Search*, que consiste em avaliar de forma sistemática diferentes combinações de hiperparâmetros a partir de um conjunto pré-definido de valores possíveis, com o objetivo de identificar a configuração que proporciona o melhor desempenho do modelo. Embora essa abordagem possa contribuir para a melhoria dos resultados, ela também aumenta significativamente o custo computacional do processo de treinamento, uma vez que múltiplos modelos precisam ser treinados e avaliados.

Neste trabalho, a aplicação do *Grid Search* não foi realizada devido às limitações de recursos computacionais disponíveis, sendo os hiperparâmetros definidos a partir de experimentos preliminares e recomendações da literatura. Dessa forma, é possível que a utilização de estratégias mais extensivas de otimização de hiperparâmetros, como o *Grid Search*, possa contribuir para melhorias adicionais no desempenho dos modelos avaliados.

5 Conclusão

Este trabalho teve como objetivo implementar e avaliar arquiteturas de redes neurais profundas para a classificação de tumores cerebrais em imagens de ressonância magnética. Para isso, foram analisados os modelos pré-treinados ResNet50, DenseNet201 e EfficientNetB0, empregando a técnica de *Transfer Learning* com *fine-tuning*. A metodologia proposta incluiu etapas de pré-processamento das imagens, como redimensionamento, conversão para escala de cinza e aplicação de filtragem bilateral.

Os resultados obtidos indicaram que a arquitetura EfficientNetB0 apresentou o melhor desempenho geral entre os modelos avaliados, alcançando uma acurácia de 75,63 e um F1-score de 73,00. De modo geral, os modelos demonstraram boa capacidade de discriminação entre as classes, apresentando valores elevados de área sob a curva ROC (AUC superiores a 0,91). Mesmo com resultados inferiores aos apresentados no estudo de referência, as arquiteturas avaliadas neste trabalho demonstraram capacidade consistente de classificação das imagens, apresentando valores elevados de especificidade e área sob a curva ROC. Esses resultados indicam que os modelos são capazes de distinguir adequadamente exames com e sem presença de tumor, embora ainda haja espaço para melhorias no sistema.

Como contribuição, este trabalho reforça o potencial do uso de CNNs no apoio à análise de imagens médicas, especialmente na tarefa de classificação de tumores cerebrais. Além disso, a utilização de *Transfer Learning* mostrou-se uma estratégia eficiente para acelerar o treinamento dos modelos e permitir a obtenção de resultados satisfatórios mesmo com conjuntos de dados relativamente limitados.

Algumas limitações foram observadas no decorrer do trabalho: a sensibilidade moderada dos modelos e a limitação na diversidade dos conjuntos de dados utilizados. Esses fatores podem impactar a capacidade de generalização dos modelos quando aplicados em cenários clínicos mais amplos, indicando a necessidade de investigações adicionais e do uso de bases de dados mais abrangentes.

Como trabalhos futuros, é válida a exploração de técnicas adicionais para aumento de dados, incluindo abordagens baseadas na geração de imagens sintéticas, com o objetivo de ampliar a variabilidade do conjunto de treinamento. Também é possível investigar estratégias mais sistemáticas de otimização de hiperparâmetros, como o uso de Grid Search, além da avaliação de novas arquiteturas de redes neurais e possíveis melhorias nas etapas de pré-processamento das imagens.

Por fim, destaca-se que o desenvolvimento de métodos automatizados para análise de imagens médicas possui grande relevância na área da saúde. Ferramentas baseadas

em aprendizado profundo podem auxiliar profissionais no processo de diagnóstico, contribuindo para análises mais rápidas e precisas e apoiando a tomada de decisão clínica.

Referências

ALZUBAIDI, L.; ZHANG, J.; HUMAIDI, A. J.; AL-DUJAILI, A.; DUAN, Y.; AL-SHAMMA, O.; SANTAMARíA, J.; FADHEL, M. A.; AL-AMIDIE, M.; FARHAN, L. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. **Journal of Big Data**, v. 8, p. 53, 2021. Citado 5 vezes nas páginas 4, 11, 12, 13 e 18.

BHUVAJI, S.; KADAM, A.; BHUMKAR, P.; DEDGE, S.; KANCHAN, S. **Brain Tumor Classification (MRI) Dataset**. 2020. Kaggle. Disponível em: <https://www.kaggle.com/dsv/1183165>. Acesso em: 05 mar. 2026. Citado 2 vezes nas páginas 16 e 25.

BOW, S. T. **Pattern Recognition and Image Preprocessing**. 2. ed. New York: Marcel Dekker, 2002. Citado na página 10.

CHAKRABARTY, N. **Brain MRI Images for Brain Tumor Detection (Br35H Dataset)**. 2017. Kaggle. Disponível em: <https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection>. Acesso em: 05 mar. 2026. Citado 2 vezes nas páginas 17 e 26.

CHENG, J. **Brain Tumor Dataset**. 2017. Figshare. Disponível em: <https://figshare.com/articles/dataset/brain_tumor_dataset/1512427>. Acesso em: 05 mar. 2026. Citado 2 vezes nas páginas 17 e 26.

Cè, M.; IRMICI, G.; FOSCHINI, C.; DANESINI, G. M.; FALSITTA, L. V.; SERIO, M. L.; FONTANA, A.; MARTINENGHI, C.; OLIVA, G.; CELLINA, M. Artificial intelligence in brain tumor imaging: A step toward personalized medicine. **Current Oncology**, v. 30, n. 3, p. 2673–2701, 2023. Open access under CC BY license. Disponível em: <https://www.mdpi.com/1718-7729/30/3/2673>. Citado na página 8.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. **IEEE Conference on Computer Vision and Pattern Recognition**, 2009. Citado na página 20.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. Citado na página 21.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. [S.l.]: Springer, 2009. Citado na página 23.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2016. p. 770–778. Citado 2 vezes nas páginas 13 e 20.

HUANG, G.; LIU, Z.; MAATEN, L. V. D.; WEINBERGER, K. Q. Densely connected convolutional networks. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2017. p. 2261–2269. Citado 2 vezes nas páginas 13 e 20.

KAYA, Y.; GURSOY, E. A mobilenet-based cnn model with a novel fine-tuning mechanism for covid-19 infection detection. **Soft Computing**, v. 27, n. 9, p. 5521–5535, 2023. Citado na página 13.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **Proceedings of the 14th International Joint Conference on Artificial Intelligence**. [S.l.: s.n.], 1995. p. 1137–1143. Citado na página 23.

MOREIRA, A.; SANTOS, S.; OLIVEIRA, M.; JúNIOR, I. P.; ASSIS, D. Classificação de tumores cerebrais em imagens de ressonância magnética. In: **Anais do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde**. [S.l.]: Sociedade Brasileira de Computação, 2024. p. 424–435. Citado 9 vezes nas páginas 8, 9, 13, 14, 15, 20, 26, 27 e 28.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of Big Data**, v. 6, n. 60, 2019. Citado na página 10.

TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. In: **Proceedings of the International Conference on Machine Learning**. [S.l.: s.n.], 2019. p. 6105–6114. Citado 2 vezes nas páginas 13 e 20.

TOMASI, C.; MANDUCHI, R. Bilateral filtering for gray and color images. In: **Proceedings of the Sixth International Conference on Computer Vision**. [S.l.: s.n.], 1998. p. 839–846. Citado na página 10.

YAPICI RUKIYE KARAKIS, K. G. M. M. Improving brain tumor classification with deep learning using synthetic data. In: . l: [s.n.], 2022. Citado 3 vezes nas páginas 13, 14 e 15.