

# DETECÇÃO E CLASSIFICAÇÃO DE TUMORES CEREBRAIS EM IMAGENS DE RESSONÂNCIA MAGNÉTICA UTILIZANDO REDES NEURAIS CONVOLUCIONAIS

Um estudo com MobileNetV2 e validação multiclasse em base pública (Kaggle)

Leon Pereira Pinto Fagundes – Desenvolvimento de Software Multiplataforma, FATEC Votorantim, Votorantim, Brasil. E-mail: leon.fagundes@fatec.sp.gov.br

Waldemar Bonventi Júnior, Doutor – Linha de pesquisa: Aprendizagem de Máquina, FATEC Votorantim, Votorantim, Brasil. E-mail: waldemar.junior@fatec.sp.gov.br, ORCID: 0000-0003-0720-499X.

## Resumo

A detecção precoce de tumores cerebrais impacta diretamente conduta terapêutica e prognóstico. Este trabalho apresenta uma abordagem baseada em Redes Neurais Convolucionais (CNNs) para classificação automática de imagens de ressonância magnética (RM) cerebral em quatro classes: glioma, meningioma, pituitária e sem tumor. Utilizou-se o Brain Tumor MRI Dataset (Kaggle), com pré-processamento padronizado e ajuste de uma MobileNetV2 pré-treinada no ImageNet. O treinamento ocorreu em duas fases: (i) congelamento do extrator convolucional, e (ii) fine-tuning a partir da camada 100. Avaliou-se desempenho com acurácia, métricas por classe (Precisão, Recall, F1) e AUC-ROC, além de matriz de confusão e análise de overfitting. No conjunto de validação ( $n=1.404$ ), o modelo alcançou acurácia global de 90,60% e AUC-ROC macro  $\geq 0,97$ , com desempenho mais baixo em meningioma ( $F1=0,8245$ ) e elevado em sem tumor ( $F1=0,9608$ ). Os resultados demonstram a viabilidade de CNNs leves para apoio ao diagnóstico por imagem, com potencial para integração em fluxos clínicos, desde que respeitados limites de generalização e validação externa.

Palavras-chave: CNN; tumores cerebrais; imagens médicas; deep learning; ressonância magnética; classificação multiclasse.

## 1 Introdução

Os tumores cerebrais constituem relevante problema de saúde pública no sistema nervoso central (SNC), com impacto funcional e socioeconômico expressivo. A classificação correta (ex.: gliomas, meningiomas e adenomas hipofisários) orienta manejo e prognóstico. A ressonância magnética (RM) é o padrão de excelência para avaliação não invasiva de lesões intracranianas, porém a interpretação depende da experiência do radiologista e pode sofrer variações interobservador.

Nas últimas décadas, o aprendizado profundo (deep learning) — notadamente as redes neurais convolucionais (CNNs) — mostrou desempenho de ponta em tarefas de visão computacional, incluindo a análise de imagens médicas. Estudos prévios evidenciam acurácia

elevada em detecção e classificação de tumores cerebrais por RM. Persistem, contudo, desafios: bases restritas, cenários binários (tumor versus não tumor), heterogeneidade de aquisição e a dependência de hiperparâmetros/arquiteturas.

Este trabalho propõe, portanto, uma classificação multiclasse (4 classes) em RM cerebral empregando MobileNetV2 com transfer learning e fine-tuning, avaliando rigorosamente métricas globais e por classe, além de analisar overfitting, distribuição de confiança e curvas ROC. Contribuímos com um pipeline reprodutível, leve e compatível com implantação em cenários com recursos computacionais moderados.

## 2 Fundamentação Teórica / Revisão da Literatura

### 2.1 Tumores Cerebrais

Gliomas abrangem neoplasias originadas de células gliais, variando do baixo ao alto grau; meningiomas derivam das meninges e, em geral, apresentam crescimento mais lento; tumores hipofisários (pituitários) acometem a hipófise e podem cursar com alterações hormonais. A correta estratificação influencia planejamento cirúrgico, radioterapia/quimioterapia e seguimento.

### 2.2 Inteligência Artificial Aplicada à Medicina

A IA na radiologia evoluiu de aprendizado de máquina tradicional para deep learning, capaz de aprender representações hierárquicas diretamente dos dados. Em RM, problemas clássicos incluem segmentação, detecção e classificação; a disponibilidade de bases públicas e o avanço de GPUs viabilizaram experimentos reprodutíveis e comparáveis.

### 2.3 Redes Neurais Convolucionais (CNN)

CNNs utilizam camadas convolucionais para extrair padrões espaciais com compartilhamento de pesos, o que as torna eficientes e invariantes a translações locais. Arquiteturas modernas (p.ex., MobileNetV2 utilizada nesse projeto) incorporam blocos inverted residual e linear bottlenecks, permitindo alto desempenho com baixo custo computacional, atributo crucial para aplicações clínicas e dispositivos na borda.

### 2.4 Trabalhos Correlatos

Estudos com CNNs para tumores cerebrais reportam acurácia elevada em cenários de duas ou três classes, frequentemente com transfer learning a partir de redes pré-treinadas (ImageNet) e ajustes finos conforme domínio médico. Persistem lacunas de generalização (diferenças de scanners, protocolos, populações) e interpretação do modelo (ex.: saliency/Grad-CAM), motivando avaliações multiclasse e análises de erro mais detalhadas.

Espaço para inserir trabalhos específicos (autor, ano, base utilizada, classes, métricas, limitações) conforme demanda do orientador:

– [Autor(es), Ano]. Título. Periódico/Evento. Principais achados: [...].

– [Autor(es), Ano]. Título. Periódico/Evento. Principais achados: [...].

### 3 Metodologia

#### 3.1 Base de Dados

Utilizou-se o Brain Tumor MRI Dataset (Kaggle), composto por imagens provenientes de Figshare, SARTAJ e Br35H, organizadas em quatro classes (glioma, meningioma, pituitary, sem tumor). O repositório disponibiliza pastas já estratificadas em treino e teste que foram unificados durante o treinamento e teste da rede, a fim de retirar qualquer possíveis vieses impostos pelo autor do dataset na separação das imagens em pastas. No presente estudo, configurou-se `validation_split = 0,20` no gerador de imagens, resultando em 5.619 amostras para treino e 1.404 para validação.

A distribuição por classe na validação foi:

- Glioma (n=324; 23,08%)
- Meningioma (n=329; 23,43%)
- Pituitária (n=351; 25,00%).
- Sem tumor (n=400; 28,49%)

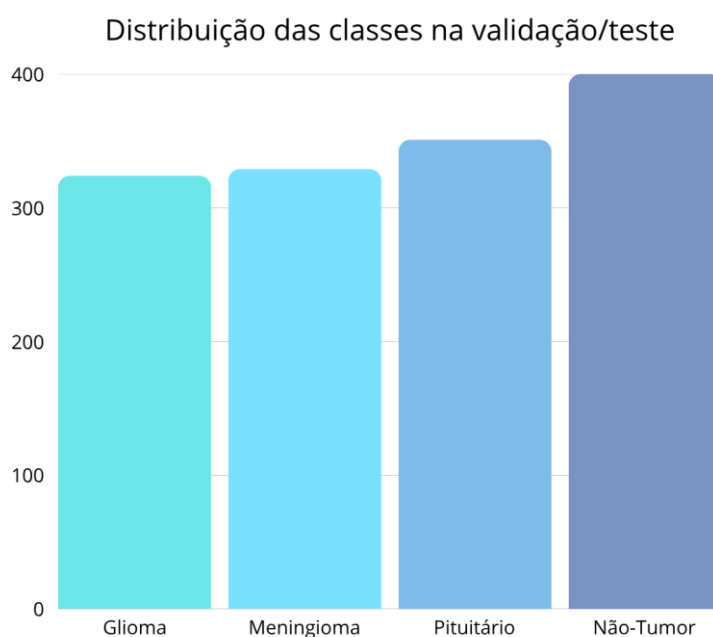


Figura 1 - Distribuição de classes na validação/teste

#### 3.2 Pré-processamento das Imagens

- Conversão para RGB (padronização para entrada do backbone pré-treinado).
- Redimensionamento para 128×128 pixels.
- Normalização dos valores de pixel para [0,1].
- Agrupamento em batches de 8 imagens.
- Espaço para inserir eventual remoção de artefatos, equalização, recorte centrado ou estratégias adicionais, que nesse estudo não precisaram ser aplicadas.

#### 3.3 Arquitetura do Modelo

Backbone MobileNetV2 (ImageNet), include\_top=False. Cabeçalho denso com:

- GlobalAveragePooling2D
- Dense(128, ReLU)
- Dense(4, Softmax)
- Parâmetros totais: 2.422.468 (treináveis: 2.025.924; não-treináveis: 396.544).

### 3.4 Ambiente de Desenvolvimento

- Linguagem: Python
- Biblioteca: TensorFlow com Keras
- Treinamento sem utilização de GPU

Dependências principais:

- tensorflow
- numpy
- pandas
- scikit-learn
- matplotlib
- pyyaml
- pathlib
- python-multipart
- pillow

### 3.5 Processo de Treinamento e Validação

Fase 1 – Treinamento inicial (backbone congelado):

- Épocas: 30 (executadas: 17 com Early Stopping)
- Adam,  $lr = 5 \times 10^{-4}$
- EarlyStopping: monitor='val\_loss', patience=10, restore\_best\_weights=True
- ReduceLROnPlateau: factor=0,5, patience=3

Fase 2 – Fine-tuning (a partir da camada 100):

- Épocas: 20 (executadas: 10)
- Adam,  $lr = 1 \times 10^{-5}$
- Primeiras 100 camadas mantidas congeladas; demais liberadas para ajuste.
- Data augmentation: rescale=1./255

### 3.6 Avaliação de Desempenho

- Métricas globais: Acurácia e Loss.
- Métricas por classe: Precisão ( $P = VP/(VP+FP)$ ), Recall/Sensibilidade ( $R = VP/(VP+FN)$ ) e  $F1 = 2PR/(P+R)$ .
- Curvas ROC (one-vs-rest) e AUC-ROC por classe.

- Matriz de Confusão absoluta e normalizada.
- Análises adicionais: evolução de acurácia/loss, gap treino–validação (indicador de overfitting) e distribuição de confiança das predições.

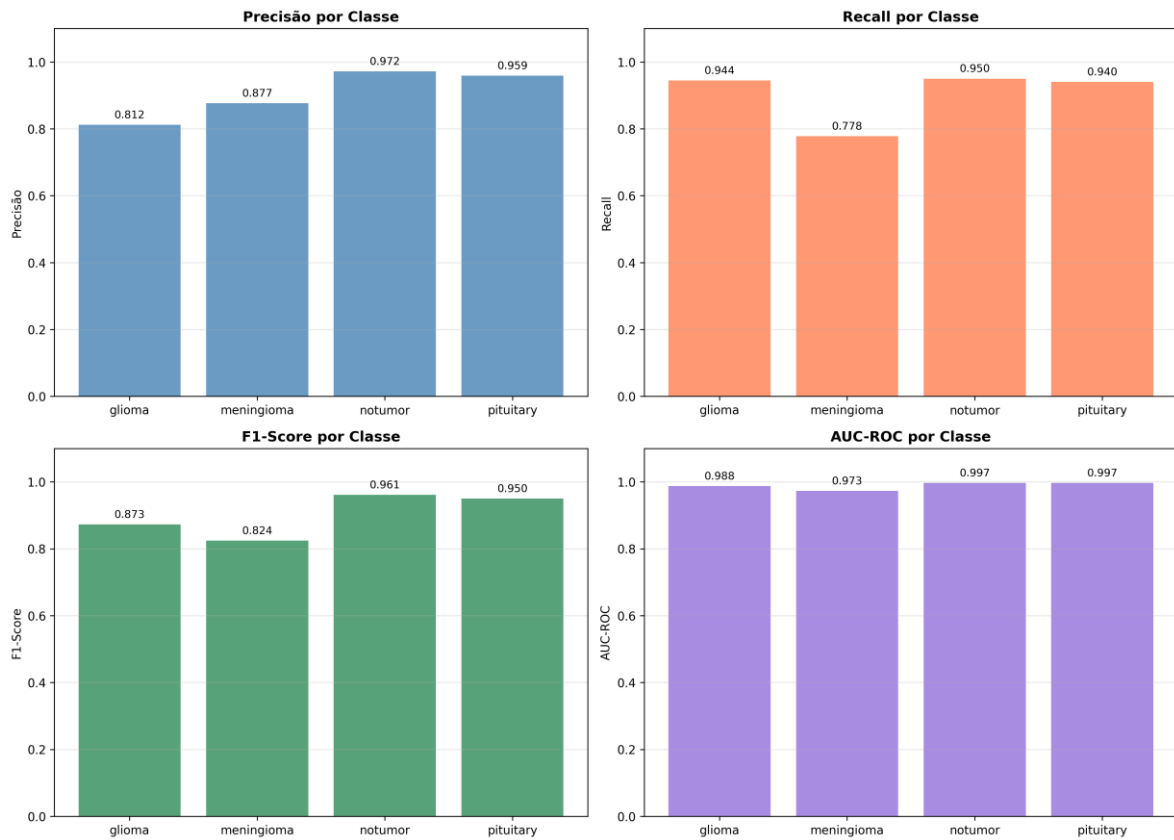


Figura 3 - Métricas por classe

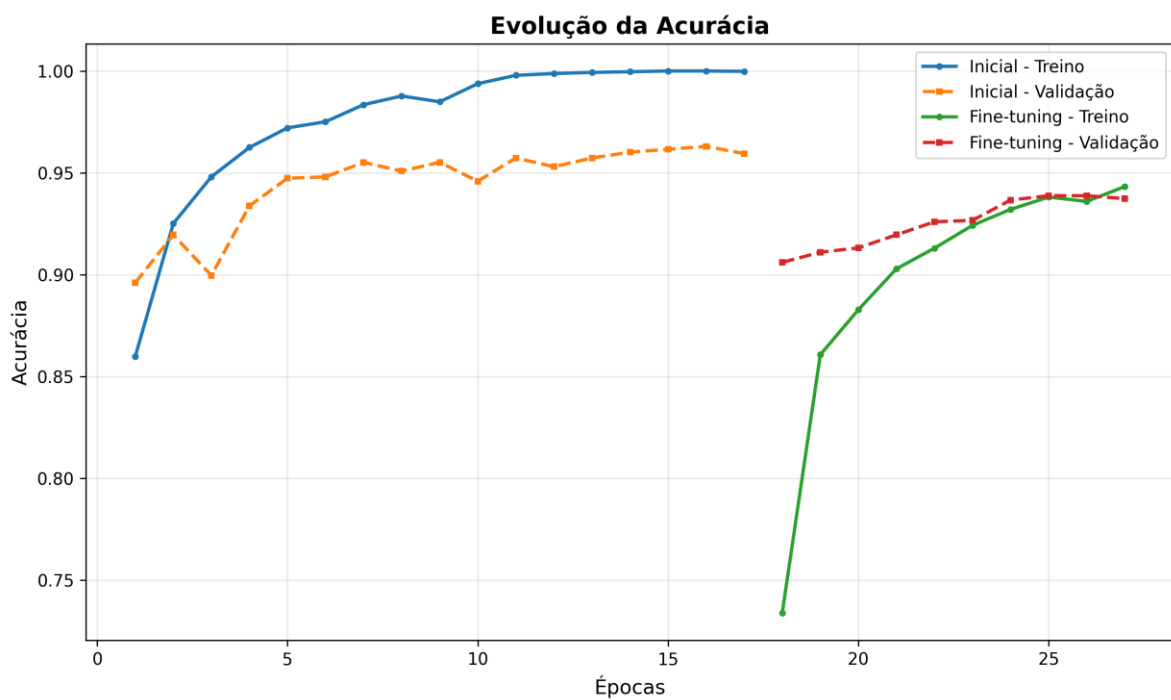


Figura 4 - Evolução da acurácia

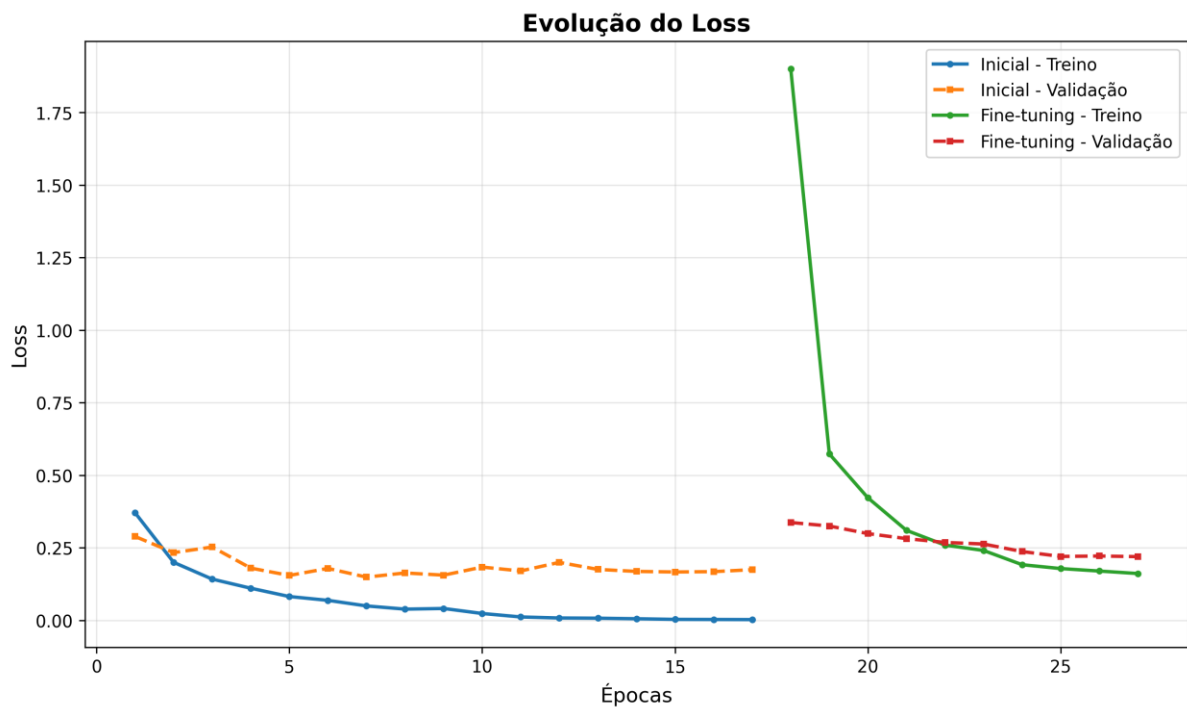


Figura 5 - Evolução da perda (loss) durante o treinamento da rede e fine-tuning

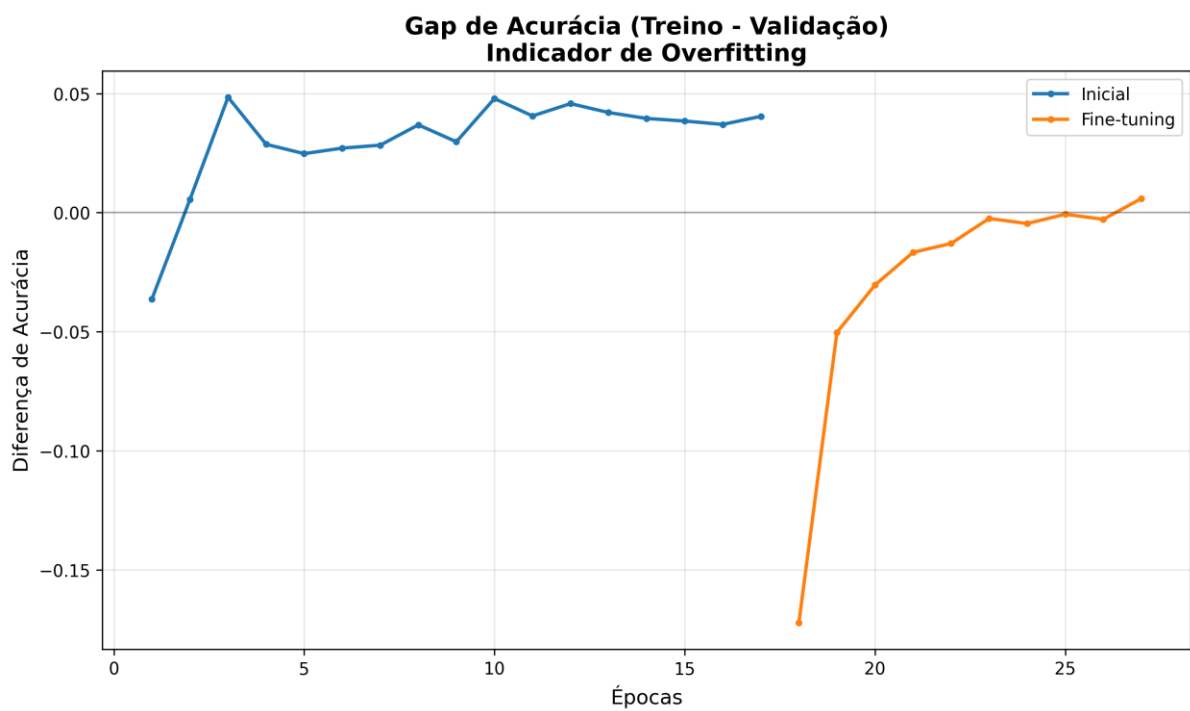


Figura 8 – Indicador de overfitting

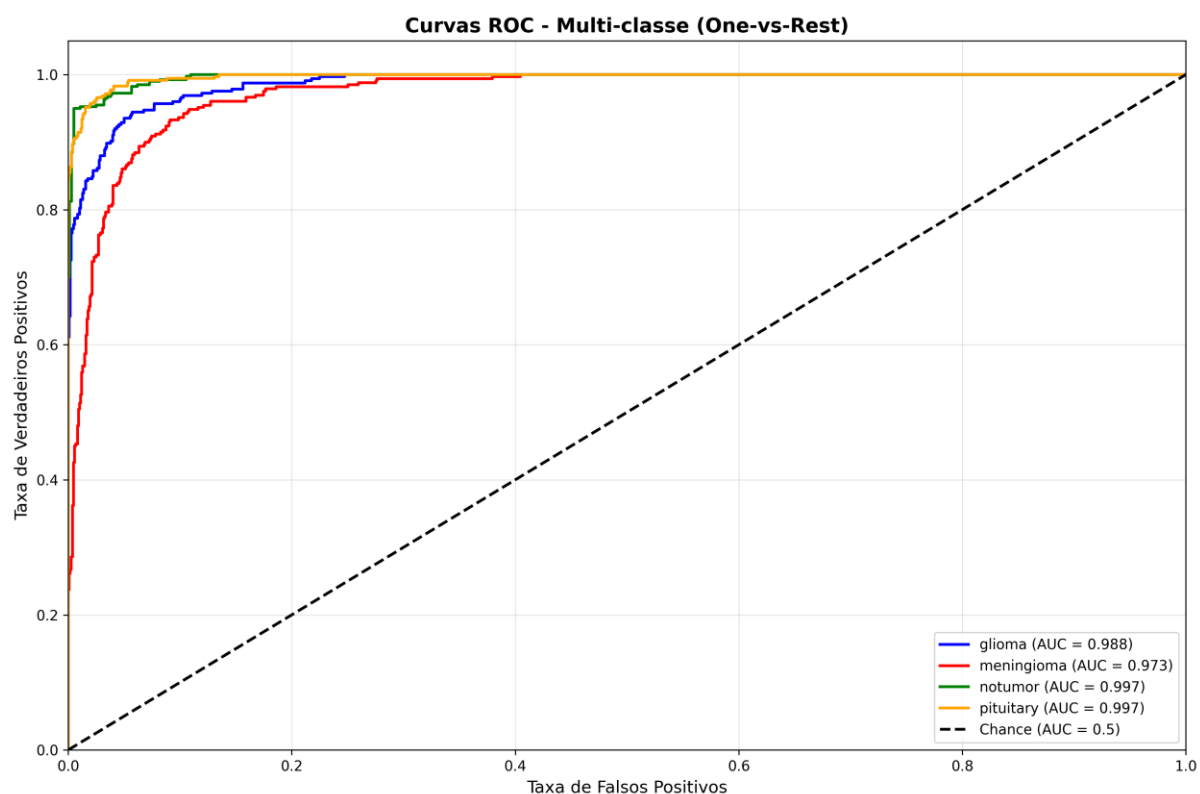


Figura 9 – Curva de AUC taxa de falsos positivos

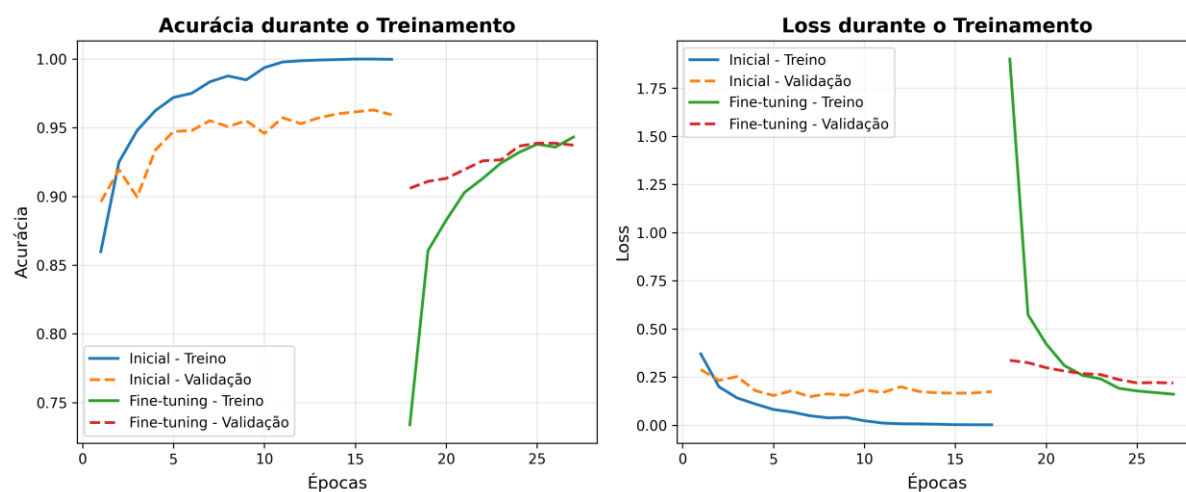


Figura 10 - Acurácia e loss durante o treinamento

## 4 Resultados e Discussão

### 4.1 Resultados Quantitativos

- Desempenho global (validação, n=1.404):
- Acurácia: 0,9060 (90,6%)
- Loss: 0,3371 (33,7%)
- Precisão macro: 0,9049 | Recall macro: 0,9032 | F1 macro: 0,9020

- Precisão weighted: 0,9095 | Recall weighted: 0,9060 | F1 weighted: 0,9058
- Métricas por classe na fase de teste:

	P	AUC	F1	Acertos (%)
Glioma	81,17%	98,77%	87,30%	306/324 (94,44%)
Meningioma	87,67%	97,29%	82,45%	256/329 (77,81%)
Pituitária	95,93%	99,69%	94,96%	330/351 (94,02%)
Sem tumor	97,19%	99,66%	96,08%	380/400 (97,19%)

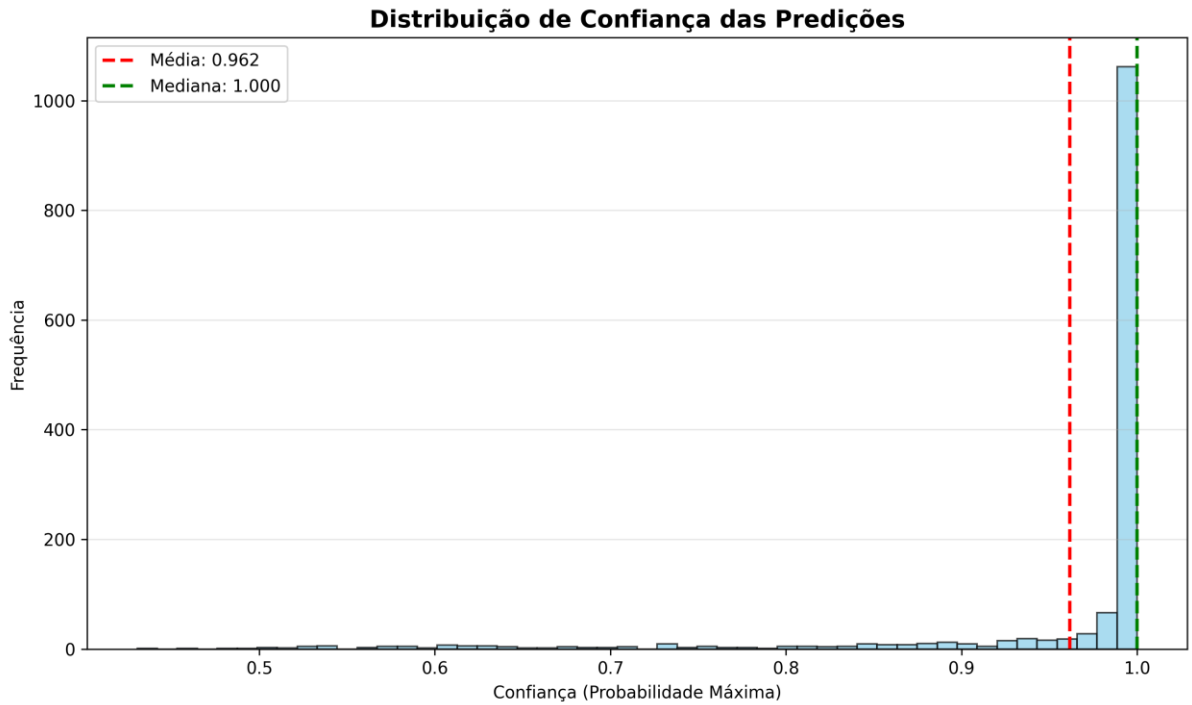


Figura X -Distribuição de confiança nas predições

Matriz de Confusão (validação):

Maiores confusões: meningioma ↔ glioma (51 erros de meningioma para glioma; 17 no sentido oposto).

Classes sem tumor e pituitária apresentaram menor taxa de confusão cruzada.

Histórico de treino:

Fase 1: Acurácia de validação subiu de 0,8960 para ≈0,9594; menor val\_loss em 0,1487.

Fase 2: Acurácia de validação oscilou em ≈0,906–0,9387; val\_loss final 0,2194.

O gap treino–validação manteve-se contido no fine-tuning (linha levemente negativa a ~0%), sugerindo controle de overfitting, conforme a figura X.

Curvas ROC:  $AUC \geq 0,97$  em todas as classes; sem tumor e pituitária atingem ≈0,997.



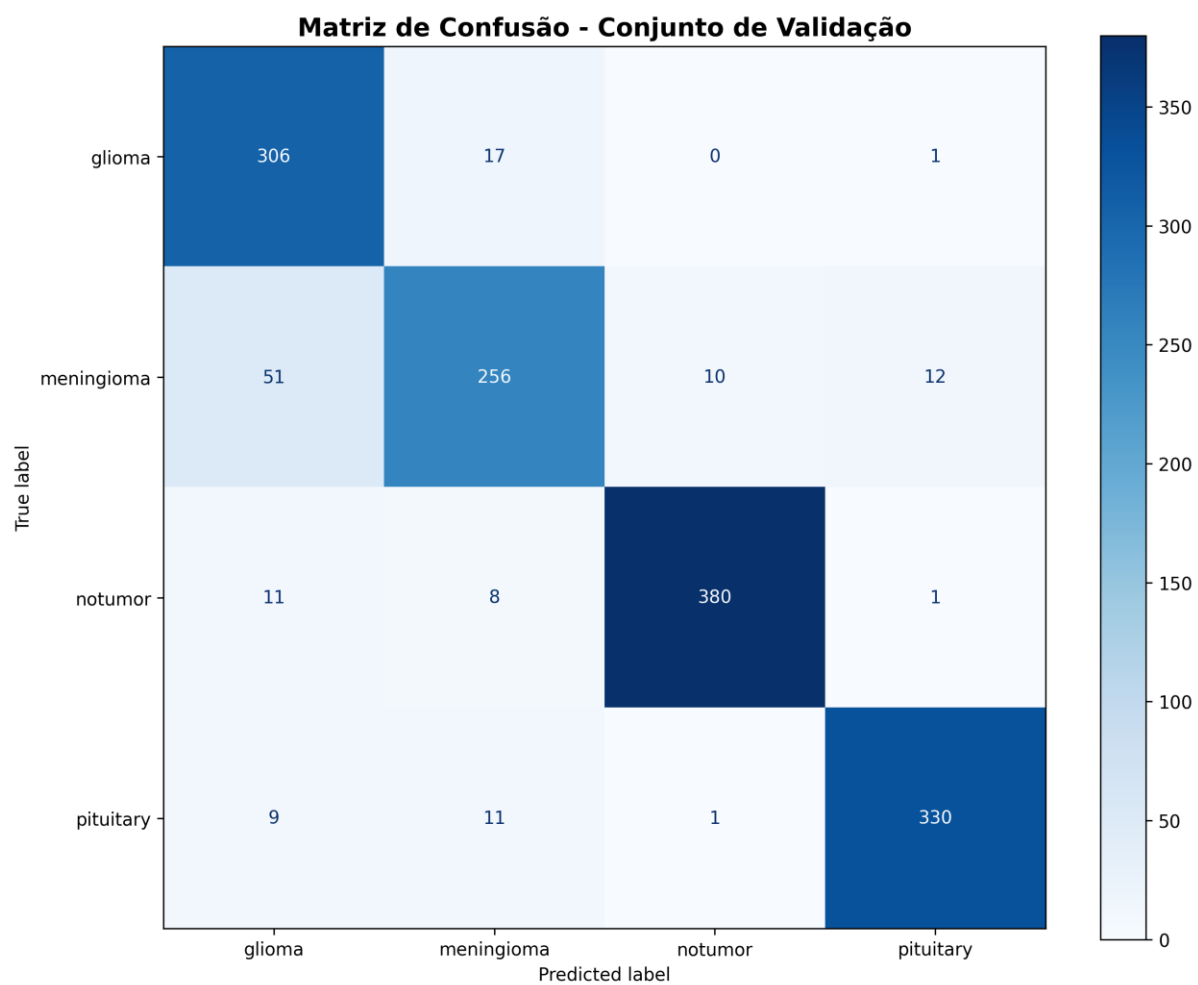


Figura 6 – Matriz de confusão do conjunto de validação

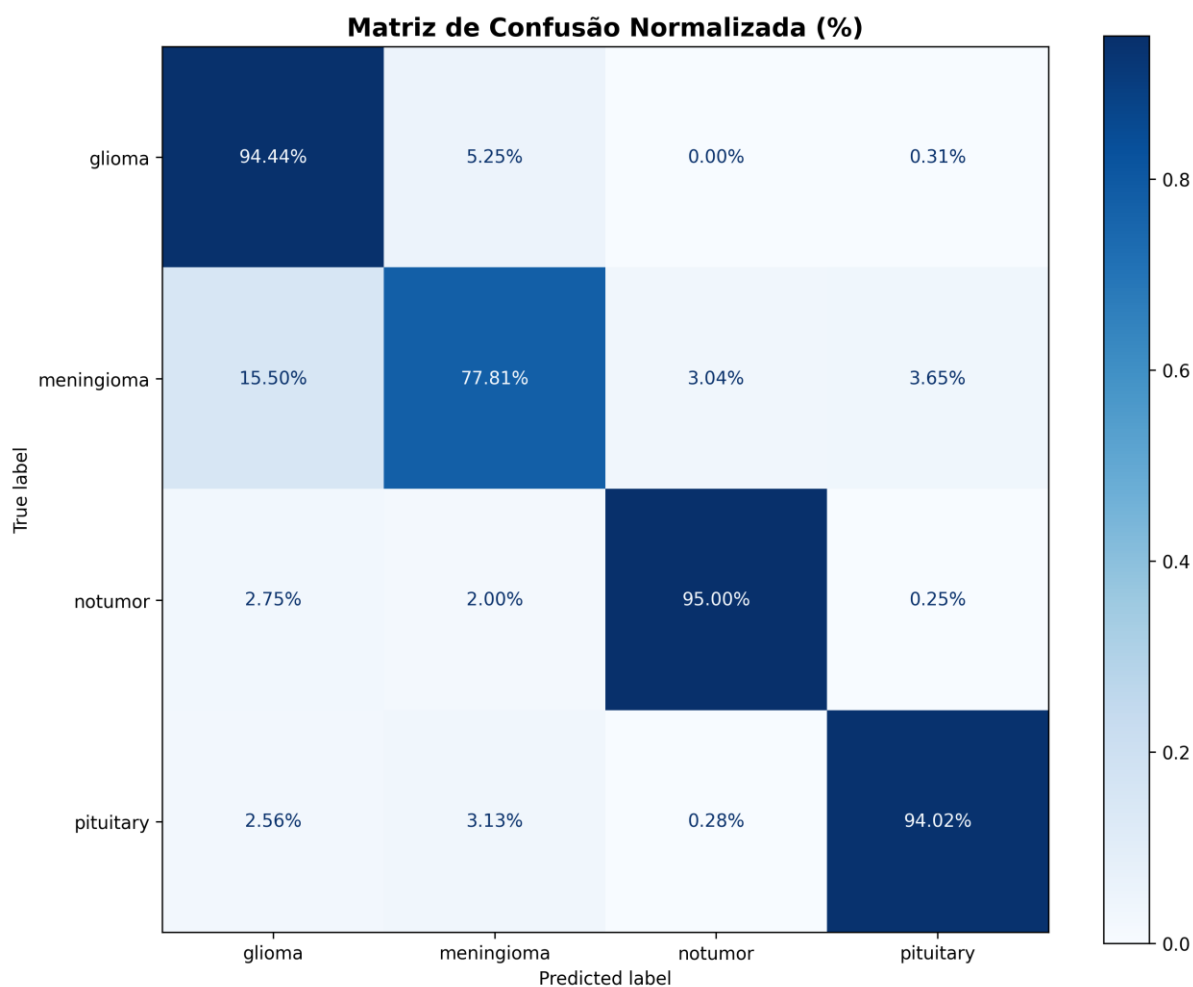


Figura 7 – Matriz de confusão normalizada do conjunto de validação

## 4.2 Resultados Qualitativos

A inspeção manual das amostras corretamente classificadas sugere que padrões texturais e morfológicos característicos de pituitária e sem tumor foram prontamente reconhecidos, possivelmente por maior consistência intra-classe. Nos erros glioma ↔ meningioma, observou-se sobreposição de características (localização extra-axial vs. intra-axial, edema/peritumoral, realce pós-contraste) que podem se atenuar com imagens axialmente heterogêneas e cortes variáveis.

## 4.3 Discussão dos Resultados

Os achados comprovam que uma CNN leve (MobileNetV2) — com entrada  $128 \times 128$  e cabeçalho raso — atinge AUCs próximas de 1,0 e acurácia acima de 90% na validação, sem sinais evidentes de overfitting no fine-tuning. O desempenho inferior em meningioma reflete: (i) heterogeneidade de aparência (graus, calcificações, variantes), (ii) variações de protocolo (CE-T1 predominante) e (iii) possíveis desbalanceamentos sutis nas fontes. Estratégias para

mitigar incluem: aumento de dados dirigido (rotação/zoom/elastic), focal loss, class weights, mixup/cutmix, e refinamento da entrada (p.ex.,  $224 \times 224$ ), ou incorporação de mapas de atenção (Grad-CAM) para regularizar regiões salientes.

Do ponto de vista prático,  $AUC \geq 0,97$  e  $F1 \geq 0,94$  para sem tumor e pituitária sugerem bom potencial para triagem assistida e priorização de laudos. Entretanto, para glioma/meningioma — classes clinicamente críticas — recomenda-se:

- Validação externa em bases multi-institucionais;
- Análise por sequência (T1, T2, FLAIR, CE-T1);
- Explainability sistemática para apoio à interpretabilidade;
- Avaliação clínica prospectiva (tempo/impacto no fluxo de trabalho).

Ameaças à validade: partição por imagens (e não por paciente) pode inflar a estimativa se houver múltiplos cortes do mesmo exame; heterogeneidade de scanners e protocolos; metadados limitados.

Aspectos éticos: o modelo não substitui o radiologista; destina-se a apoio à decisão.

## 5 Conclusão

Este estudo mostrou que uma CNN leve (MobileNetV2) com transfer learning e fine-tuning é capaz de classificar tumores cerebrais em RM em quatro classes com acurácia de 90,60% e  $AUC-ROC \geq 0,97$  em todas as classes, com destaque para sem tumor e pituitária. Identificou-se maior confusão entre glioma e meningioma, apontando caminhos de melhoria (aumento de dados direcionado, balanceamento, maior resolução e múltiplas sequências). Como trabalhos futuros, propõe-se validação externa multi-centro, explicabilidade por Grad-CAM, integração de metadados clínicos e avaliação prospectiva do impacto no fluxo de laudos.

## Referências

CHENG, J. et al. Enhanced Performance of Brain Tumor Classification via Transfer Learning. [Periódico/Evento], 2017. DOI: [inserir].

DEEPAK, S.; AMEER, P. M. Brain Tumor Classification Using Deep CNNs. [Periódico], 2019. DOI: [inserir].

EL-DAHSHAN, E.-S. A. et al. Computer-aided diagnosis of brain tumor. [Periódico], 2010. DOI: [inserir].

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. MIT Press, 2016.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. Nature, 521, p. 436–444, 2015.

LOUIS, D. N. et al. WHO Classification of Tumors of the Central Nervous System (5th ed.). IARC/WHO, 2021.

NICKPARVAR, M. Brain Tumor MRI Dataset (Kaggle), 2020. Disponível em: [link do dataset].

REHMAN, A. et al. A Survey of MRI-based Brain Tumor Classification. [Periódico], 2020. DOI: [inserir].

SULTAN, H. H. et al. Multi-class Brain Tumor Classification via CNNs. [Periódico], 2019. DOI: [inserir].

Inclua referências extras de trabalhos correlatos citados na seção 2.4, e o manual/guia do MobileNetV2 (Sandler et al., 2018), além de documentação Keras/TensorFlow se julgar pertinente:

SANDLER, M. et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. CVPR, 2018. DOI: [inserir].

Apêndices (se houver necessidade)

Apêndice A – Código de Treinamento (versão reprodutível)

Inclua aqui o script completo com gerenciamento de semente aleatória, versões de bibliotecas e instruções de execução (CLI/Notebook), além de referência a caminhos e nomes de arquivos salvos:

`./model/modelo_mobilenetv2_128x128_finetuned.keras`

Diretório com figuras: `./docs/` ou `/mnt/data/` (substitua conforme organização final)

Observação: garanta remoção de caminhos pessoais e dados sensíveis.

Apêndice B – Protocolo de Pré-processamento Detalhado

Especificar conversões, normalizações, redimensionamento, e (se aplicável) políticas de data augmentation.

Anexos (se houver necessidade)

Anexo I – Figuras de Desempenho

Figura 1. Evolução da Acurácia (arquivo: `accuracy_evolution.png`).

Figura 2. Evolução do Loss (`loss_evolution.png`).

Figura 3. Matriz de Confusão (`confusion_matrix.png`).

Figura 4. Matriz de Confusão Normalizada (`confusion_matrix_normalized.png`).

Figura 5. Métricas por Classe (`metrics_by_class.png`).

Figura 6. Gap de Acurácia e Overfitting (`overfitting_analysis.png`).

Figura 7. Distribuição de Confiança (prediction\_confidence\_distribution.png).

Figura 8. Curvas ROC (roc\_curves.png).

Figura 9. Painel do Histórico de Treinamento (training\_history.png).