UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

EDERSON NAVES FERNANDES GONÇALVES JÚNIOR

IDENTIFICAÇÃO DE TUMOR CEREBRAL UTILIZANDO ESTRATÉGIA DE EXTRATORES *ZERO-SHOT*

Ouro Preto, MG
2025

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

EDERSON NAVES FERNANDES GONÇALVES JÚNIOR

IDENTIFICAÇÃO DE TUMOR CEREBRAL UTILIZANDO ESTRATÉGIA DE EXTRATORES *ZERO-SHOT*

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Pedro Henrique Lopes Silva
Coorientador: Guilherme Augusto Lopes Silva

Ouro Preto, MG
2025

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIENCIAS EXATAS E BIOLOGICAS
DEPARTAMENTO DE COMPUTACAO

FOLHA DE APROVAÇÃO

Ederson Naves Fernandes Gonçalves Júnior

Identificação de Tumor Cerebral utilizando estratégia de extratores Zero-Shot

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 12 de Março de 2025

Membros da banca

Doutor Pedro Henrique Lopes Silva (Orientador) - Universidade Federal de Ouro Preto
Mestre Guilherme Augusto Lopes Silva (Coorientador) - Doutorando do PPGCC-UFOP
Doutor Rodrigo César Pedrosa Silva (Examinador) - Universidade Federal de Ouro Preto
Bacharel Pablo Martins Coelho (Examinador) - Mestrando do PPGCC-UFOP

Pedro Henrique Lopes Silva, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 12/03/2025

Documento assinado eletronicamente por **Pedro Henrique Lopes Silva**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 13/03/2025, às 15:38, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 , informando o código verificador **0872492** e o código CRC **660441D9**.

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35402-163
Telefone: 3135591692 - www.ufop.br

Dedico esse trabalho a Deus, pois sem Ele nada posso. Dedico também a minha família, amigos e orientadores, por todo o apoio fornecido a mim durante minha vida.

Agradecimentos

Gostaria de dedicar esse espaço para agradecer a todas as pessoas que auxiliaram na construção desse projeto.

Primeiramente a Deus, por proporcionar saúde, paz, família e a oportunidade de fazer meus sonhos se tornarem realidade.

Aos meus pais: Milenne e Ederson, por me apoiarem em minhas decisões e me auxiliarem a ser o homem que sou hoje. A meu padrasto, Robson, por fazer parte da minha vida desde que me entendo por gente. As minhas irmãs Emanuelle e Ana Laura, por todas as risadas e momentos de irmãos. A meus avós, Antônio, Pedro, Gininha e Tutuca, por momentos que foram mais como pais e mães para mim do que avós.

A meus amigos de faculdade: Arthur Negrão, Fernanda Coelho, Guilherme Salim, Lucas Ferreira, Mateus Filipe e Rômulo Junio, por diversas risadas, perrengues e estudos durante a graduação.

Por meus amigos de Congonhas que aguentaram muita reclamação minha ao longo do curso. Todos vocês estão em meu coração.

E por fim, agradeço aos Professores Pedro Silva e Guilherme Silva, por toda orientação e amizade, desde o início da faculdade e que perdurará pros dias futuros.

“*Pluvius at sumus*”. Somos pó e sombra
Clare (2010).

Resumo

A identificação de tumores no cérebro é um desafio na interseção entre medicina e tecnologia. Devido à complexidade e variedade desses tumores, os métodos tradicionais de diagnóstico geralmente exigem exames especializados e detalhados realizados por profissionais da saúde. Entretanto, esses métodos ainda apresentam limitações na detecção precisa e precoce de diferentes tipos de tumores. Apesar de técnicas baseadas em *Convolutional Neural Network* serem amplamente utilizadas para esse fim, elas tendem a demandar elevado custo computacional e longos períodos de treinamento. Por isso, este estudo propõe o uso de extratores de características como paradigma *zero-shot*, com o intuito de desenvolver modelos mais eficientes para a detecção de tumores cerebrais, reduzindo a necessidade de treinamento adicional e melhorando a eficiência do processo. Utilizando a estratégia de um extrator *zero-shot*, foi alcançado uma acurácia de 99,15% usando o modelo DINOV2.

Palavras-chave: Tumor Cerebral; *Zero-Shot*; *Convolutional Neural Network*; *Visual Transformers*; *Distillation with No Labels version 2*.

Abstract

The identification of brain tumors is a significant challenge at the intersection of medicine and technology. Due to the complexity and diversity of these tumors, traditional diagnostic methods often require specialized and detailed examinations conducted by healthcare professionals. However, these methods still have limitations in the accurate and early detection of different types of tumors. While *Convolutional Neural Network*-based techniques are widely used for this purpose, they tend to require high computational costs and long training periods. Therefore, this study proposes the use of feature extractors in a zero-shot paradigm, aiming to develop more efficient models for brain tumor detection, minimizing the need for additional training and improving the process's efficiency. Using the strategy of zero-shot extractor, an accuracy of 99.15% was achieved with the DINOV2 model.

Keywords: Brain Tumors; Zero-shot; *Convolutional Neural Network*; *Visual Transformers*, *Distillation with No Labels version 2*.

Lista de Ilustrações

Lista de Tabelas

Lista de Abreviaturas e Siglas

BM3D *Block-matching and 3D filtering*. 23

BMI Brain MRI Images for Brain Tumor Detection. viii, 25, 27, 30, 33–36

BTDM Brain Tumor Detection MRI. viii, 25, 27, 30–32, 34–36

CNN *Convolutional Neural Network*. v, vi, 1–4, 11–13, 15, 22, 24, 27, 28, 31, 32, 37

DINOV2 *Distillation with No Labels version 2*. v–vii, x, 2–5, 14–16, 23, 31–37

FN Falso Negativo. 17, 18

FP Falso Positivo. 17

IA Inteligência Artificial. x, 1, 3, 5, 7

MLP *Multi-Layer Perceptron*. vii, 9, 10, 14

RELU *Rectified Linear Unit* . 11, 24

RM Ressonância Magnética. 1, 3, 6, 21, 22, 24, 25

TC Tomografia Computadorizada. 1, 6

ViT *Visual Transformers*. v–vii, x, 2–5, 13, 14, 22, 23, 28, 29, 31, 32

VN Verdadeiro Negativo. 17

VP Verdadeiro Positivo. 17

Sumário

1 Introdução

Os tumores cerebrais constituem um grupo diversificado de neoplasias — crescimentos anormais de células — que podem acometer qualquer região do sistema nervoso central, apresentando ampla variação em termos de agressividade e prognóstico (BONDY et al., 2008). A detecção precisa e precoce dessas formações tumorais é fundamental, pois impacta diretamente as opções de tratamento e as chances de recuperação do paciente (BONDY et al., 2008). Nesse cenário, os avanços tecnológicos em Inteligência Artificial (IA) e *Machine Learning* têm desempenhado um papel crucial ao aprimorar significativamente a precisão e a velocidade dos diagnósticos (KHAN et al., 2023).

Nos últimos anos, a busca por métodos mais precisos e eficientes comparado a inspeção visual e análise histopatológica para a identificação e classificação de tumores cerebrais ganhou destaque (HAVAEI et al., 2017). Esse progresso foi impulsionado por melhorias nas tecnologias de imagem médica (SILVA et al., 2020), como a Ressonância Magnética (RM) e a Tomografia Computadorizada (TC), além do desenvolvimento de algoritmos de *Deep Learning* (KUFEL et al., 2023). Essas novas abordagens têm o potencial de fornecer diagnósticos mais rápidos e precisos, superando as limitações dos métodos tradicionais baseados na inspeção visual e na análise histopatológica (KUFEL et al., 2023).

Nesse contexto, a RM tem se consolidado como uma fonte fundamental de dados, pois oferece imagens detalhadas do cérebro, essenciais para a detecção de tumores. Contudo, a análise manual dessas imagens pode ser complexa e demorada. Para superar essa limitação, as *Convolutional Neural Network*s (CNNs), uma técnica de *Deep Learning*, têm se destacado (SHABBIR; NAZIR et al., 2023; ULLAH et al., 2023).

As CNNs são capazes de processar automaticamente grandes volumes de imagens de RM, identificando características sutis e padrões que, muitas vezes, não são detectáveis por métodos tradicionais (LIU; PU; SUN, 2021). Essa capacidade de análise detalhada e automatizada não apenas melhora a precisão dos diagnósticos, como também aumenta a velocidade em comparação com a percepção humana (LIU; PU; SUN, 2021). O estudo de Ullah et al. (2023) demonstra que os modelos baseados em CNNs conseguem detectar e classificar diferentes tipos de tumores cerebrais com alta acurácia, se tornando comparável à de especialistas humanos. Além disso, a aplicação dessas tecnologias pode reduzir significativamente o tempo necessário para o diagnóstico, permitindo intervenções mais rápidas e aumentando as chances de recuperação dos pacientes (SILVA, 2018). Dessa forma, a integração das CNNs na neuro-oncologia não só aprimora a precisão diagnóstica, como também representa um passo importante em direção à medicina personalizada, onde o tratamento é adaptado às necessidades específicas de cada paciente (SILVA, 2018).

Os benefícios do uso de CNNs na identificação de tumores cerebrais vão além da precisão diagnóstica. A automação proporcionada por essas tecnologias pode aliviar a carga de trabalho dos profissionais de saúde, permitindo que eles se concentrem em casos mais complexos e em interações diretas com os pacientes (STRONG, 2016). Além disso, os modelos de *Machine Learning* podem ser continuamente aprimorados com novos dados, aumentando sua eficácia ao longo do tempo (STRONG, 2016). Outra vantagem significativa desses sistemas é a capacidade de identificar padrões nas imagens que podem passar despercebidos pelo olho humano, oferecendo *insights* valiosos para o diagnóstico precoce e o planejamento do tratamento (SILVA, 2018).

No entanto, o treinamento desses modelos pode ser extremamente oneroso, chegando a ser inviável em certos contextos. Além de exigir grandes bases de dados com diversas imagens de tumores cerebrais, o treinamento geralmente demanda recursos computacionais significativos, o que pode ser um obstáculo para instituições que não possuem esses recursos (POURPANAH et al., 2022).

Os *Visual Transformers* (ViT) emergiram como uma abordagem promissora para tarefas de visão computacional, trazendo inovações que diferem das CNNs. Baseados no conceito de *self-attention*, os ViTs processam imagens dividindo-as em pequenos *patches* e tratam cada um como uma sequência de *tokens*, similar ao processamento em linguagens naturais (DOSOVITSKIY et al., 2020a). Essa abordagem permite que o modelo aprenda relações globais entre diferentes partes da imagem, o que pode ser crucial para identificar padrões complexos e contextuais em tarefas como a detecção de tumores cerebrais.

O estudo de Dosovitskiy et al. (2021) demonstrou que os ViTs podem alcançar resultados comparáveis ou superiores às CNNs, especialmente em cenários com grande quantidade de dados. Além disso, a capacidade dos ViTs de generalizar melhor para novas tarefas os torna particularmente úteis em abordagens como o paradigma *Zero-Shot*, onde o modelo deve ser eficiente em contextos previamente não treinados (CHEN et al., 2024). Essas características não apenas reduzem a necessidade de treinamento extenso, mas também tendem a tornar os ViTs uma ferramenta poderosa para diagnósticos médicos, permitindo análises detalhadas e integradas que poderiam complementar as observações humanas e oferecer diagnósticos mais rápidos e precisos.

Diante desses desafios, este trabalho foca na aplicação de modelos baseados em ViT (como o DINOV2 e o VT-FPN) e CNNs como extratores de características dentro do paradigma *Zero-Shot*. A principal hipótese é que, por já possuírem conhecimento prévio adquirido em bases de dados não diretamente relacionadas ao problema em questão, esses modelos consigam acoplar uma camada classificadora eficiente e de rápido treinamento para detectar a presença de tumores cerebrais. Dessa forma, elimina-se a necessidade de utilizar grandes bases de dados de tumores cerebrais durante o treinamento, além de reduzir os custos computacionais, já que o bloco de extração de características não precisa de ajustes nos pesos, sendo necessário apenas modificar a camada de classificação. Além disso, este trabalho realiza o treinamento de CNNs

com *fine-tuning*, empregando as arquiteturas *EfficientNet*-B0 (TAN; LE, 2019) e *EfficientNet*-V2 L. A intenção é comparar o desempenho dessas redes com as abordagens baseadas em ViT no contexto *Zero-Shot*, considerando métricas como acurácia, precisão e tempo de treinamento. Ademais, foi utilizado uma estratégia de *cross-dataset*, que consiste em treinar o modelo em um conjunto de dados e avaliá-lo em outro conjunto distinto, mas contendo as mesmas classes, com o objetivo de avaliar a capacidade do modelo de generalizar os dados entre duas bases distintas com as mesmas classes.

Seguindo essa proposta, o presente trabalho obteve uma acurácia de 99,15% em uma avaliação cruzada com cinco divisões, com desvio padrão inferior a 0,2%, utilizando a estratégia do DINOV2 aliada à técnica de *Data Augmentation*. Além disso, em comparação com a *EfficientNet*-B0 com *fine-tuning*, a abordagem *Zero-Shot* utilizando o DINOV2 apresentou desempenho superior em todas as métricas avaliadas, com vantagem superior a 1,5%. Em comparação ao trabalho de Ullah et al. (2023), embora a acurácia seja 0,68% menor, o presente estudo destaca-se pela redução significativa no custo computacional tanto no treinamento quanto na inferência, facilitando a adoção dessa abordagem em diversas aplicações na área da saúde.

1.1 Justificativa

A detecção precoce de tumores cerebrais é fundamental para aumentar as taxas de sucesso dos tratamentos e, consequentemente, melhorar a qualidade de vida dos pacientes. No entanto, essa tarefa enfrenta diversos desafios, como a necessidade de grandes volumes de dados rotulados e a limitação de recursos computacionais, especialmente em regiões menos favorecidas. Nesse cenário, as aplicações de IA podem antecipar a detecção de tumores, oferecendo maior possibilidade de cura e melhorando o prognóstico dos pacientes. A criação de um modelo que seja ao mesmo tempo simples e eficiente torna-se indispensável para apoiar médicos ao redor do mundo, principalmente em áreas com recursos limitados.

O uso de estratégias *zero-shot* simplifica o processo de treinamento de modelos para a classificação de tumores, uma vez que essas abordagens atuam como extratores de características, permitindo que apenas uma rede neural simples seja treinada para a classificação final. Isso reduz drasticamente os requisitos de treinamento adicional, sendo particularmente vantajoso em contextos com restrições computacionais e escassez de dados.

1.2 Objetivos

O objetivo principal desse trabalho é comparar a eficácia de estratégias *zero-shot* em relação à CNNs *fine-tuned* convencionais, com o objetivo de verificar sua utilização para a classificação de tumores cerebrais em imagens de Ressonância Magnética (RM). Como objetivos específicos têm-se:

- Implementar uma estratégia *zero-shot* com o ViT.
- Implementar uma estratégia *zero-shot* com o *Distillation with No Labels version 2* (DINOV2).
- Comparar o desempenho do ViT e do DINOV2 com modelos tradicionais baseados em CNN, como a *ResNet50* (SHABBIR; NAZIR et al., 2023) e *TumorDetNet* (ULLAH et al., 2023).
- Implementar estratégias de CNNs não convencionais, como a *EfficientNet*-B0 e a *EfficientNet*-V2 L.
- Avaliar estratégia de *cross-dataset* entre base de dados distintas.
- Analisar o custo computacional necessário para treinamento e inferência dos modelos.
- Avaliar o uso de estratégia de aumento de dados (*Data Augmentation*).

1.3 Organização do Trabalho

Este trabalho está organizado da seguinte forma: no Capítulo 2, é apresentada a revisão bibliográfica, onde são discutidos os trabalhos relacionados e a base teórica necessária para o entendimento do estudo; no Capítulo 3, são descritas a metodologia adotada, a configuração dos experimentos e as técnicas utilizadas; o Capítulo 4 apresenta os resultados obtidos a partir dos experimentos, acompanhados de uma análise detalhada; e, por fim, no Capítulo 5, são discutidas as conclusões derivadas dos resultados e propostas sugestões para trabalhos futuros.

2 Revisão Bibliográfica

Esse capítulo tem como objetivo trazer uma revisão bibliográfica sobre o presente trabalho. Na Seção 2.1, serão abordados assuntos essenciais para a fundamentação deste trabalho, como Redes Neurais, Tumores Cerebrais e *Machine Learning*. Em seguida, na Seção 2.2, serão apresentados os trabalhos relacionados, com o intuito de fornecer uma visão sobre a evolução dessa área.

2.1 Fundamentação Teórica

Esta seção tem como objetivo apresentar os conceitos fundamentais utilizados ao longo deste trabalho, como Tumores Cerebrais na Seção 2.1.1, Inteligência Artificial (IA) e *Machine Learning* na Seção 2.1.2, *Redes Neurais* dentro da Seção 2.1.3, Redes Neurais Convolucionais na Seção 2.1.4, *ViT* na Seção 2.1.5, *DINOV2* na Seção 2.1.6, *K-Fold* dentro da Seção 2.1.7, métricas de avaliação na Seção 2.1.8, *Data Augmentation* na Seção 2.1.9 e *Cross-Dataset* na Seção 2.1.10.

2.1.1 Tumor Cerebral

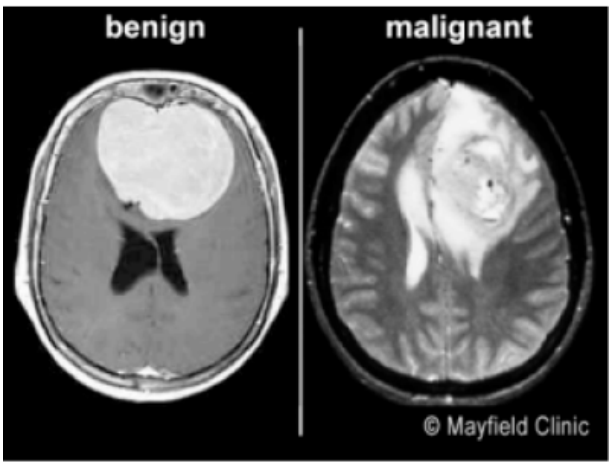
Dentre os diversos problemas de saúde que afetam o ser humano, o surgimento de tumores cerebrais figura como uma das principais preocupações (BONDY et al., 2008). Essas neoplasias são massas anormais de células que se desenvolvem no cérebro. Essas células crescem de maneira descontrolada e podem infiltrar-se nos tecidos circundantes, impactando as funções cerebrais normais. Apesar dos avanços nos tratamentos, que incluem cirurgia, radioterapia e/ou quimioterapia, ou mesmo a combinação dessas abordagens, ainda não se pode garantir a completa recuperação do paciente. Esses tumores podem ser classificados em dois tipos principais: benignos e malignos, o qual, independentemente do tipo que seja, afetam de forma negativa o corpo humano (CHARLES et al., 2011). Tais tumores, podem ser descritos da seguinte forma:

- Tumores Benignos: Geralmente menos agressivos, crescem mais lentamente e tendem a possuir bordas definidas, não se espalhando para outras regiões do corpo. No entanto, mesmo benignos, podem ocasionar sintomas graves, dependendo de sua localização no cérebro (CHARLES et al., 2011).

- Tumores Malignos: Conhecidos como câncer cerebral, são mais agressivos e propensos a se disseminarem para outras áreas do cérebro ou do corpo. Crescem rapidamente e apresentam maior probabilidade de causar danos sérios (CHARLES et al., 2011).

A Figura 2.1 abaixo mostra, respectivamente, um tumor benigno e um tumor maligno.

Figura 2.1 – Imagem de um tumor benigno e um maligno.



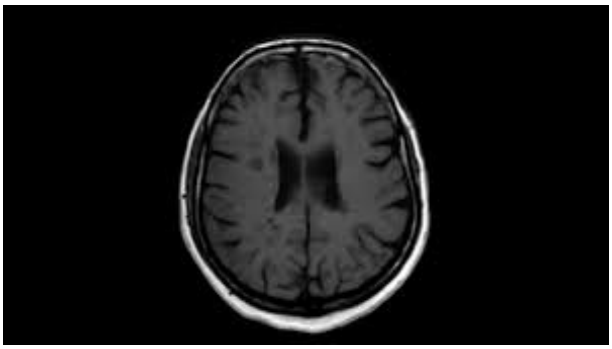
Fonte: (Marina Baeta, 2020).

Os sintomas dos tumores cerebrais variam de acordo com o tamanho, localização e tipo do tumor, podendo incluir dor de cabeça persistente, alterações visuais, dificuldades de equilíbrio, convulsões, mudanças na personalidade e problemas de memória (PARK; PARK, 2022).

O diagnóstico dos tumores cerebrais geralmente envolve exames de imagem, como a Ressonância Magnética (RM), que consiste em uma técnica avançada de imagem utilizada para diagnosticar e acompanhar tumores cerebrais. Sua aplicação é fundamental dentro deste estudo, pois permite a obtenção de imagens detalhadas e precisas do cérebro, auxiliando na identificação e caracterização dessas massas anormais de células, ao contrário de outras técnicas de imagem, como a Tomografia Computadorizada (TC), por proporcionar um contraste mais nítido entre os diferentes tecidos cerebrais (AZEVEDO, 2023).

Na Figura 2.2 abaixo é apresentado um exemplo de uma imagem de um cérebro utilizando ressonância magnética.

Figura 2.2 – Exemplo de uma imagem de um cérebro humano utilizando ressonância magnética.



Fonte: (Navoneel Chakrabarty, 2019).

O tratamento dos tumores cerebrais pode incluir cirurgia para a remoção do tumor, radioterapia para destruir as células cancerígenas e quimioterapia para impedir o crescimento tumoral. Em alguns casos, é necessária a combinação dessas terapias, dependendo do estágio e

da natureza do tumor (CHARLES et al., 2011).

Apesar dos avanços na medicina, o tratamento de tumores cerebrais pode ser desafiador e frequentemente não garante a cura completa. Por isso, a pesquisa contínua, incluindo o emprego de técnicas avançadas de Inteligência Artificial na classificação de tumores cerebrais, é crucial para melhorar as opções de tratamento e o prognóstico dos pacientes afetados (SOARES et al., 2023).

2.1.2 IA e *Machine Learning*

O surgimento da Inteligência Artificial (IA) remonta ao século XX, com os primeiros passos dados por pioneiros como Alan Turing e John McCarthy. No entanto, foi durante a década de 1950 que o termo "Inteligência Artificial" foi cunhado e o campo começou a se consolidar como uma disciplina acadêmica. Nessa época, pesquisadores começaram a explorar a ideia de criar máquinas capazes de simular processos cognitivos humanos, como raciocínio, aprendizado e resolução de problemas. Durante as décadas seguintes, houve avanços significativos em áreas como lógica simbólica, redes neurais artificiais e sistemas especialistas, impulsionando o desenvolvimento da IA. Nos últimos anos, com o advento de grandes conjuntos de dados e o aumento da capacidade computacional, percebe-se um renascimento da IA, com avanços notáveis em áreas como aprendizado de máquina, processamento de linguagem natural e visão computacional, tornando a IA uma força transformadora em diversos setores da sociedade (BUCHANAN, 2006).

No contexto do funcionamento de um sistema de Inteligência Artificial, sua operação tem como uma das vertentes a combinação de dados digitais com algoritmos de *Machine Learning*. Essa integração permite que o sistema analise e interprete padrões e informações, levando ao seu aprendizado automático (SAVEGNAGO et al., 2024).

Machine Learning se concentra no desenvolvimento de algoritmos e modelos computacionais capazes de aprender a partir de dados e realizar tarefas específicas sem a necessidade de instruções explícitas. Os algoritmos de *machine learning* analisam e interpretam padrões e informações, o que leva ao seu aprendizado automático. Em vez de serem programados explicitamente para realizar uma determinada tarefa, os sistemas de *Machine Learning* são treinados usando conjuntos de dados para aprender padrões e fazer previsões ou tomar decisões (SOORI; AREZOO; DASTRES, 2023). Ademais, a eficácia da IA também está intrinsecamente ligada à capacidade de processamento, que se refere à habilidade operacional do sistema em processar tais informações (SAVEGNAGO et al., 2024).

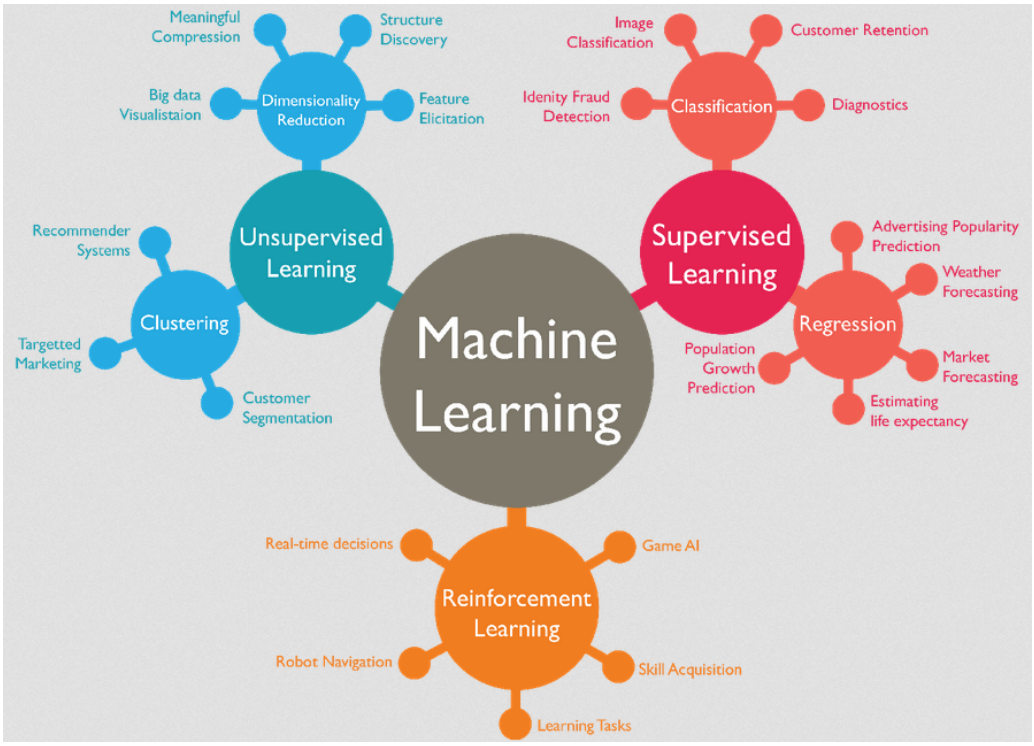
Existem várias abordagens e técnicas em *Machine Learning*, cada uma com suas próprias aplicações e métodos. Entre elas, podemos destacar o Aprendizado Supervisionado, o Aprendizado Não Supervisionado e o Aprendizado por Reforço. Esses conjuntos de dados são geralmente divididos em três partes: treinamento, validação e teste. O conjunto de treinamento

é utilizado para ajustar os pesos e parâmetros do modelo com base nos exemplos fornecidos, sendo responsável pelo aprendizado inicial. O conjunto de validação, por sua vez, é utilizado durante o treinamento para monitorar o desempenho do modelo em dados não vistos, ajudando a ajustar os hiper-parâmetros e evitar o *overfitting*, que ocorre quando o modelo se adapta tão bem ao conjunto de treinamento que perde a capacidade de generalizar para novos dados. Após o término do treinamento, o conjunto de teste, completamente independente, é utilizado para avaliar a capacidade do modelo de generalizar para novos dados, proporcionando uma métrica final de desempenho (SUNIGA, 2020).

No aprendizado supervisionado, os algoritmos são treinados em pares de entrada e rótulo, onde a entrada representa os dados de entrada e o rótulo é a saída desejada associada a essa entrada. O objetivo é aprender uma função que mapeie as entradas para os rótulos corretos. São exemplos de algoritmos de aprendizado supervisionado as árvores de decisão (GARCIA, 2003) e a regressão logística (MORAES; MOREIRA; LUIZ, 2011). No aprendizado não supervisionado, os algoritmos são treinados em conjuntos de dados sem rótulos, buscando encontrar estruturas ou padrões intrínsecos nos dados. Dentro desse algoritmo, busca-se aprender a estrutura subjacente dos dados sem a orientação de rótulos específicos. Os algoritmos *k-means* (FERNANDES et al., 2022) e *Fuzzy C-Means* (HALDAR et al., 2017) são exemplos de algoritmos não supervisionados. Já no aprendizado por reforço, os algoritmos aprendem a tomar decisões sequenciais para maximizar uma recompensa acumulada. Essa recompensa pode ser entendida como um sinal de *feedback* que o algoritmo recebe do ambiente em que está inserido. Em termos simples, o algoritmo recebe uma recompensa positiva quando toma ações que o levam mais próximo de alcançar seus objetivos e uma recompensa negativa quando toma ações que o afastam desses objetivos (IBM, 2023). *Q-Learning* (GAO; SHI; SONG, 2023) e *Policy Gradient* (DAVIES; SHERRIFF, 2014) utilizam o conceito de aprendizado por reforço . Na Figura 2.3 são apresentados exemplos de aplicações de *Machine Learning*.

Dentre as diversas técnicas de *Machine Learning*, no contexto de aprendizado supervisionado, destacam-se as redes neurais, as quais são o atual estado-da-arte para vários problemas da literatura, como a classificação de imagens médicas (NETO, 2023), reconhecimento de fala (RUNSTEIN, 1998) e processamento de linguagem natural (CASELI; FREITAS; VIOLA, 2022). A classificação de tumores é um exemplo claro de aprendizado supervisionado, onde pares de imagens e rótulos (diagnósticos) são usados para treinar o modelo. Dessa forma, é fundamental explorar como as redes neurais podem ser aplicadas neste contexto para melhorar a precisão e a eficiência dos diagnósticos. Com isto em mente, torna-se um caminho natural a sua investigação no problema de detecção de tumores cerebrais (ULLAH et al., 2023; SHABBIR; NAZIR et al., 2023; LITJENS et al., 2017; PEREIRA et al., 2019). Na Seção 2.1.3, será abordado em detalhes como essas técnicas são empregadas para enfrentar desafios específicos da detecção de tumores cerebrais.

Figura 2.3 – Ilustração de aplicações de aprendizado supervisionado, não supervisionado e por reforço de *Machine Learning*.



Fonte: (SHEWAN, 2023).

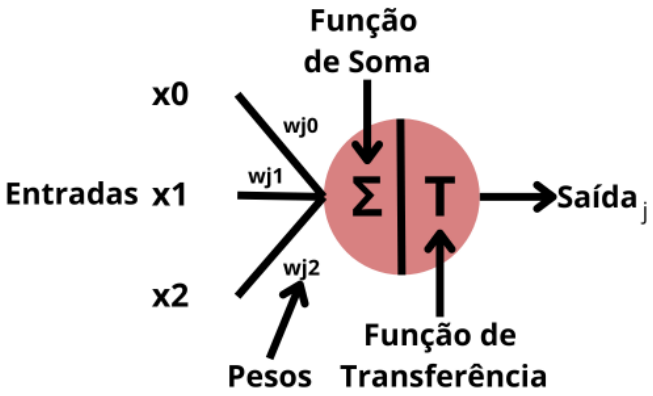
2.1.3 Redes Neurais

Uma das técnicas de *Machine Learning* de aprendizado supervisionado que apresenta resultados de estado-da-arte em diversos problemas é as Redes Neurais, que são modelos computacionais inspirados no funcionamento do cérebro humano. As redes neurais têm sido amplamente utilizadas na classificação de tumores cerebrais a partir de imagens de ressonância magnética (CASPER et al., 2023).

O princípio fundamental das redes neurais é a conexão entre neurônios artificiais, que são unidades computacionais que processam informações. Cada neurônio recebe múltiplos sinais de entrada, e cada sinal de entrada é associado a um peso, conforme visto na Figura 2.4. O neurônio então calcula a soma ponderada dessas entradas, adiciona um valor de viés (*bias*) e aplica uma função de ativação para introduzir não linearidade ao modelo (CASPER et al., 2023).

Uma rede neural é um conjunto de neurônios que estão divididos em três tipos de camadas principais: (i) de entrada, (ii) de saída e (iii) ocultas. A camada de entrada recebe os dados iniciais e transmite a informação para a(s) camada(s) oculta(s), onde ocorre a maior parte do processamento e extração de características. A camada de saída, por sua vez, produz a resposta final do modelo. Todos os neurônios de uma camada estão ligados com os neurônios da próxima. Este tipo de arquitetura é denominada de *Multi-Layer Perceptron* (MLP). Em uma MLP, todos os neurônios de uma camada estão conectados a todos os neurônios da camada subsequente, formando uma rede totalmente conectada. Essa estrutura permite que a MLP capture e aprenda padrões complexos

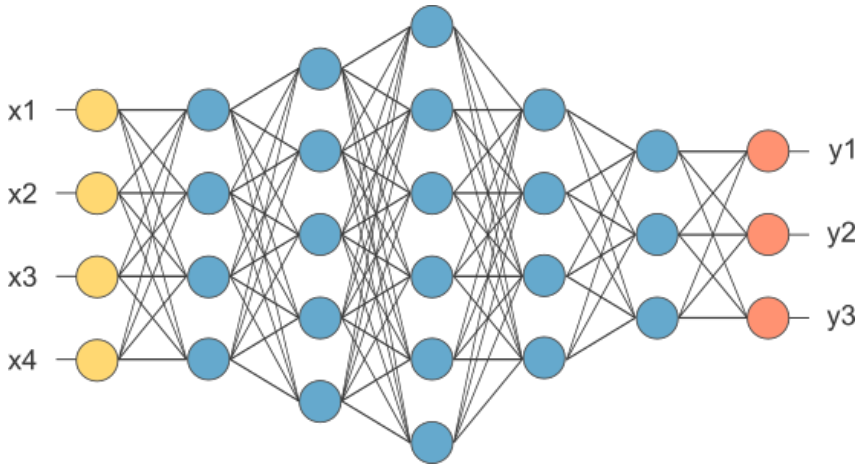
Figura 2.4 – Exemplo de um neurônio presente em uma rede neural.



Fonte: Próprio Autor.

nos dados. Essas camadas combinadas geram uma arquitetura e pode ser visto um exemplo na Figura 2.5 (TAUD; MAS, 2018).

Figura 2.5 – Exemplo de uma arquitetura de uma rede neural (MLP). Os neurônio de cor amarela representam a camada de entrada, os de cor azul estão na camada oculta, e os de cor vermelha, a camada de saída.



Fonte: (COELHO, 2017).

Uma das prioridades no treinamento de redes neurais é o ajuste dos pesos das conexões entre os neurônios. Esse processo é realizado por meio da minimização do erro da classificação da rede, com a utilização do algoritmo da descida do gradiente, que ajusta os pesos de forma iterativa para minimizar uma função de custo que quantifica a diferença entre as previsões do modelo e os valores reais dos dados de treinamento. Na classificação de tumores cerebrais, as redes neurais são usadas para "aprender" a distinguir entre diferentes tipos de tumores com base nas características das imagens de ressonância magnética (KUFEL et al., 2023).

Ao contrário de abordagens tradicionais de *Machine Learning*, que requerem a extração manual de características dos dados, as redes neurais são capazes de aprender automaticamente características relevantes diretamente dos dados. Isso é possível graças à estrutura em camadas, onde cada camada aprende representações progressivamente, desde bordas até representações

mais abstratas dos dados (KUFEL et al., 2023). Quando várias camadas de redes neurais são empilhadas e as arquiteturas são profundas, elas são reconhecidas por aprendizado em profundidade, ou *Deep Learning*.

Uma das principais necessidades de arquiteturas de *Deep Learning* são os grandes volumes de dados para adequação dos pesos do modelo ao problema tratado. A aplicação deste tipo de arquitetura é especialmente relevante em áreas como a medicina, onde conjuntos de dados de ressonância magnética de alta resolução podem conter milhares ou até milhões de imagens (CASPER et al., 2023). As Redes Neurais Convolucionais, do inglês *Convolutional Neural Network* (CNN), são um tipo comum de arquitetura de *Deep Learning* usada na análise de imagens em geral, incluindo ressonância magnética cerebral (TAJBAKHSH et al., 2020). As CNNs são projetadas para capturar padrões espaciais nas imagens, tornando-as ideais para tarefas de detecção, segmentação e classificação de tumores cerebrais (KUFEL et al., 2023).

2.1.4 Redes Neurais Convolucionais

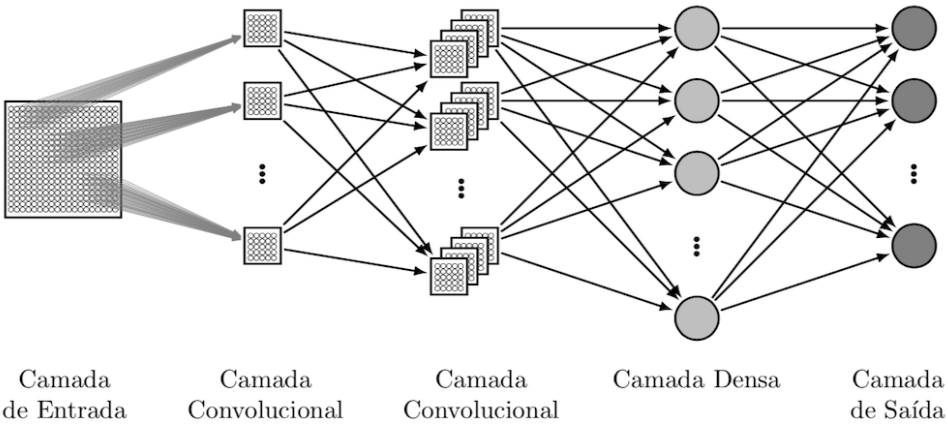
CNN é uma classe de redes neurais profundas particularmente eficazes no processamento e análise de dados com estrutura de grade, como imagens. Inspiradas na organização do córtex visual animal, as CNNs foram projetadas para reconhecer padrões espaciais e hierárquicos nas imagens, o que as torna ideais para tarefas de visão computacional, incluindo a classificação de tumores cerebrais a partir de imagens de ressonância magnética (SILVA, 2018).

A estrutura das CNNs é composta por uma série de camadas que extraem e processam características das imagens de entrada. Essas camadas incluem:

- **Camadas Convolucionais**: São responsáveis por aplicar filtros (*kernels*) às imagens de entrada para produzir mapas de características. Esses filtros aprendem a detectar padrões locais como bordas, texturas e formas (SILVA, 2018).
- **Camadas de Pooling**: Reduzem a dimensionalidade dos mapas de características, o que diminui a carga computacional. A operação de *max pooling* é a mais comum, onde o valor máximo em uma região específica do mapa de características é selecionado (SILVA, 2018).
- **Camadas de Ativação**: Introduzem não-linearidades no modelo, permitindo que a rede aprenda representações complexas. A função de ativação mais comum em CNNs é a *Rectified Linear Unit* (RELU), que aplica uma transformação não-linear (SILVA, 2018).
- **Camadas Completamente Conectadas**: Ao final da rede, essas camadas realizam a classificação com base nas características extraídas pelas camadas anteriores. Cada neurônio está conectado a todos os neurônios da camada anterior, permitindo uma combinação global das características (SILVA, 2018).

Na Figura 2.6, é apresentado um exemplo de arquitetura de uma Rede Neural Convolucional.

Figura 2.6 – Exemplo de uma arquitetura de Rede Neural Convolucional.



Fonte: (SAKURAI, 2017).

As CNNs são particularmente eficazes na análise de imagens médicas devido à sua capacidade de capturar e representar informações espaciais e contextuais das imagens. Isso é essencial para a classificação e segmentação de tumores cerebrais, onde a localização e a forma dos tumores são características cruciais para o diagnóstico preciso (LIU; PU; SUN, 2021).

O treinamento de CNNs, assim como outras redes neurais profundas, envolve a minimização de uma função de custo que quantifica a diferença entre as previsões do modelo e os valores reais dos dados de treinamento. Esse processo é realizado utilizando algoritmos de otimização, como o gradiente descendente e suas variantes. Durante o treinamento, os pesos dos filtros são ajustados iterativamente para melhorar a precisão da rede na tarefa de classificação (LIU; PU; SUN, 2021).

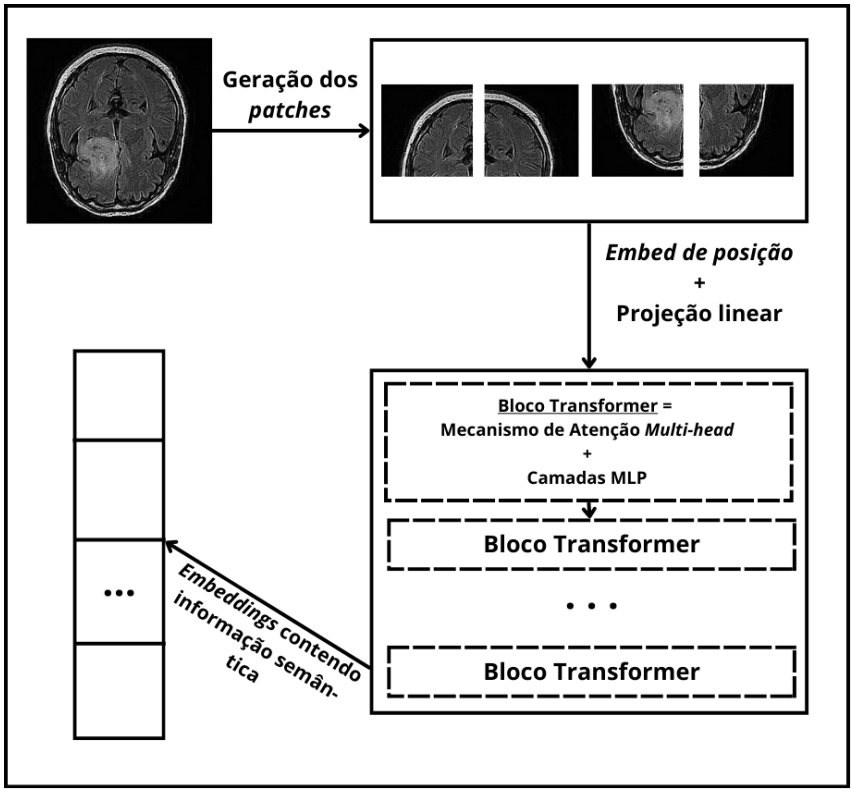
Dentre uma das possibilidades de CNNs que existem, uma das mais promissoras é a família *EfficientNet* (TAN; LE, 2019), onde esta abordagem se destaca por introduzir uma nova estratégia de escalonamento das CNNs, utilizando um método composto que ajusta simultaneamente três dimensões: profundidade da rede, largura dos filtros e resolução das imagens (TAN; LE, 2019). Essa abordagem permite que os modelos cresçam de maneira mais eficiente em comparação com o escalonamento tradicional, onde apenas uma dessas dimensões é ajustada por vez. Além disso, as variantes dessa família, *como EfficientNet*-B0 a *EfficientNet*-B7, foram desenvolvidas para equilibrar de forma cuidadosa a precisão e o custo computacional, permitindo que modelos menores e mais rápidos alcancem resultados semelhantes a arquiteturas maiores e mais lentas (TAN; LE, 2019). Assim, as redes *EfficientNet* não apenas contribuem para um melhor desempenho em tarefas de classificação de imagens, mas também viabilizam o uso em dispositivos com recursos limitados, ampliando o alcance e a aplicabilidade das CNNs em diferentes contextos (TAN; LE, 2019).

2.1.5 *Visual Transformers*

O *Visual Transformers* (ViT) é uma abordagem que utiliza o mecanismo de atenção para processar imagens. O mecanismo de atenção é uma técnica originalmente desenvolvida para modelos de linguagem natural, mas que também se mostrou eficaz no processamento de imagens (MOUTIK et al., 2023). Ele funciona destacando as partes mais relevantes de uma entrada, permitindo que o modelo foque nas informações mais importantes para a tarefa em questão. Em vez de tratar todas as partes de uma imagem da mesma maneira, o mecanismo de atenção atribui diferentes "pesos" às diversas regiões da imagem, dependendo de sua importância para a tarefa (MOUTIK et al., 2023).

O ViT processa imagens de maneira distinta das CNNs. Enquanto as CNNs aplicam filtros sobre pequenos quadrados de *pixels* para extrair características locais, o ViT divide a imagem de entrada em uma sequência de *patches*. Cada *patch* é então convertido em um vetor através de um operador linear, criando um *patch embedding*. Além disso, a posição de cada *patch* é transformada em um vetor utilizando *position encoding*. Esses dois vetores são somados e processados por vários codificadores de transformador, permitindo ao modelo capturar as relações entre diferentes partes da imagem de forma integrada (WU et al., 2020). O funcionamento básico de um ViT é apresentado na Figura 2.7.

Figura 2.7 – Fluxograma de funcionamento do *Visual Transformers* (ViT)



Fonte: Próprio Autor.

O mecanismo de atenção no ViT transforma repetidamente os vetores de representação dos *patches* de imagem, incorporando cada vez mais relações semânticas entre os *patches* ao

longo do tempo. Isso é análogo ao processamento de linguagem natural, onde os vetores de representação das palavras fluem por um transformador, incorporando relações que vão da sintaxe à semântica (DOSOVITSKIY, 2024).

A arquitetura do ViT transforma uma imagem em uma sequência de representações vetoriais, as quais capturam informações detalhadas sobre o conteúdo e a estrutura da imagem. Para utilizar essas representações em aplicações práticas, é necessário treinar uma cabeça adicional que possa interpretar essas informações de maneira eficaz, permitindo o uso do ViT em uma ampla gama de tarefas, desde classificação de imagens até a detecção de anomalias em imagens médicas (DOSOVITSKIY, 2024).

O *Visual Transformers* (ViT) foi inicialmente desenvolvido como um transformador codificador supervisionado, com o objetivo de prever o rótulo de uma imagem a partir dos seus *patches*, no qual o ViT utiliza um *token* especial *<CLS>* na entrada, que coleta e resume as informações de toda a imagem para ser usado na previsão final do rótulo. O vetor de saída correspondente a este *token* é então usado como a única entrada para o cabeçote MLP final. Este *token* especial serve como um mecanismo arquitetônico que permite ao modelo compactar todas as informações relevantes para a previsão do rótulo da imagem em um único vetor. A partir disso, o ViT consegue capturar e processar de maneira eficiente as informações contidas nas imagens, utilizando a arquitetura de atenção dos transformadores (MAURÍCIO; DOMINGUES; BERNARDINO, 2023).

Além do uso tradicional do ViT para classificação de imagens, o modelo DINOV2 pode ser aplicado para extrair características de forma não supervisionada. Ao utilizar o DINOV2, o ViT é treinado para gerar representações ricas e discriminativas das imagens, sem a necessidade de rótulos explícitos. Essas características extraídas podem ser aplicadas em uma variedade de tarefas, como segmentação, detecção de anomalias ou mesmo como entradas para outros modelos, ampliando a flexibilidade e a aplicabilidade do ViT em diferentes contextos de aprendizado de máquina.

2.1.6 *Distillation with No Labels version 2* (DINOV2)

O *Distillation with No Labels version 2* (DINOV2) é um modelo de aprendizagem auto-supervisionada que se baseia na destilação de conhecimento, sem a necessidade de rótulos explícitos. Isso permite a criação de modelos robustos e eficientes para diversas tarefas visuais (HUANG et al., 2024), além de se basear em aprendizagem contrastiva.

A destilação de conhecimento envolve a transferência de conhecimento de um modelo professor para um modelo aluno, onde o modelo aluno é treinado para imitar as saídas do modelo professor (HUANG et al., 2024). Este processo permite que o modelo aluno aprenda representações visuais de alta qualidade sem a necessidade de rótulos anotados manualmente. No contexto do DINOV2, a destilação é realizada de maneira auto-supervisionada (AYZENBERG;

GIRYES; GREENSPAN, 2024). O modelo professor gera "pseudo-rótulos" a partir dos dados não anotados, que são então utilizados para treinar o modelo aluno. Este processo é iterativo, onde o modelo aluno pode, eventualmente, se tornar o novo modelo professor, refinando ainda mais as representações aprendidas (AYZENBERG; GIRYES; GREENSPAN, 2024).

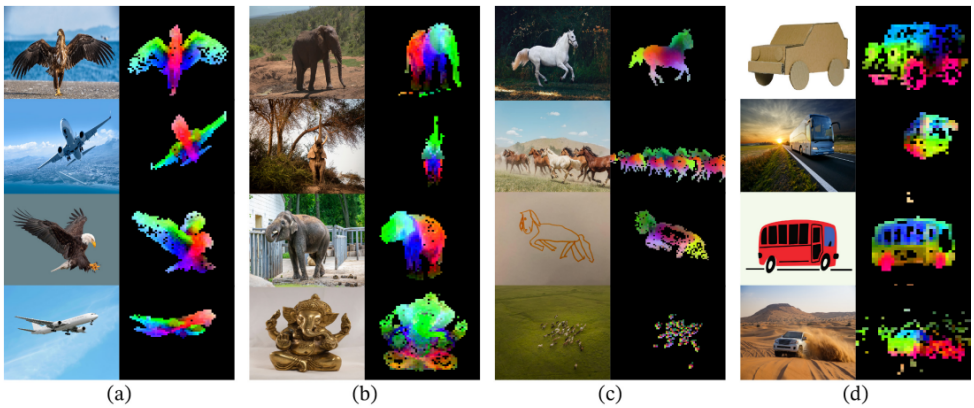
A partir disso, têm-se a aprendizagem contrastiva, que é um componente do DINOV2 que visa aprender representações discriminativas ao comparar exemplos positivos (diferentes vistas da mesma imagem) com exemplos negativos (vistas de imagens diferentes). Ao maximizar a similaridade entre exemplos positivos e minimizar a similaridade entre exemplos negativos, o modelo aprende a capturar características relevantes das imagens de maneira mais eficiente e robusta (OQUAB, 2024).

A arquitetura do DINOV2 geralmente envolve o uso de CNN e *Transformers* como base para os modelos professor e aluno. A escolha da arquitetura depende da tarefa específica e dos requisitos de desempenho (HUANG et al., 2024).

O processo de treinamento do DINOV2 pode ser descrito em várias etapas. Primeiramente, a quantidade de imagens são aumentadas de várias maneiras, como cortes, rotações e mudanças de cor, para aumentar a diversidade dos dados de entrada e melhorar a robustez do modelo (DAMM et al., 2024). Em seguida, o modelo professor processa as imagens aumentadas e gera representações de alta dimensão, que servem como pseudo-rótulos para o treinamento do modelo aluno (DAMM et al., 2024). O modelo aluno é então treinado para minimizar a diferença entre suas próprias saídas e os pseudo-rótulos fornecidos pelo modelo professor, utilizando uma função de perda de entropia cruzada ou uma variante da mesma (DAMM et al., 2024). Após várias iterações, o modelo aluno pode ser promovido a modelo professor, e o processo de geração de pseudo-rótulos e treinamento é repetido (DAMM et al., 2024).

Na Figura 2.8, segue um exemplo de uma utilização de *DINOV2* em diferentes objetos, de forma a demonstrar como o *DINOV2* consegue capturar informações sobre a imagem.

Figura 2.8 – Exemplo de utilização do *DINOV2* na representações de diferentes objetos.



Fonte: (OQUAB et al., 2023a).

O DINOV2 oferece várias vantagens em comparação com métodos supervisionados

tradicionais. Ao eliminar a necessidade de rótulos explícitos, o DINOV2 pode ser aplicado a grandes volumes de dados não anotados, reduzindo significativamente o custo e o tempo associados à anotação manual (OQUAB et al., 2023a). Além disso, modelos treinados com base no DINOV2 tendem a generalizar melhor em tarefas desconhecidas e são mais robustos a variações nos dados de entrada (OQUAB et al., 2023a). Embora o treinamento inicial possa ser computacionalmente intensivo, a capacidade de treinar com grandes volumes de dados não anotados pode levar a ganhos de eficiência a longo prazo.

O DINOV2 pode oferecer benefícios no contexto de identificação de tumores cerebrais, especialmente em cenários onde a anotação manual de dados é limitada ou inviável. A capacidade do modelo de aprender representações visuais de alta qualidade sem a necessidade de rótulos explícitos, ou com poucos dados, pode ser uma solução para enfrentar o desafio da detecção de tumores cerebrais, onde a diversidade e a complexidade das imagens tornam a anotação um processo custoso e demorado. Com o DINOV2, pode-se aplicar auto-supervisão em grandes volumes de imagens não rotuladas, o que não só reduz o tempo de preparação dos dados, mas também melhora a robustez e generalização do modelo em diferentes cenários clínicos. Ou ele pode ser usado no paradigma *zero-shot*.

2.1.7 *K-Fold*

Entre as diversas abordagens para avaliar modelos, o método *K-Fold* se destaca como uma das técnicas, especialmente quando se trabalha com conjuntos de dados limitados. A essência do método reside na divisão do conjunto de dados em K partes de tamanho igual. Durante o treinamento, o modelo é iterativamente treinado K vezes, utilizando uma parte diferente dos dados como conjunto de teste em cada iteração e as demais como conjunto de treinamento. Dessa maneira, o modelo é testado em todas as partes dos dados, garantindo uma avaliação abrangente em diferentes subconjuntos dos dados de treinamento (CUNHA, 2019).

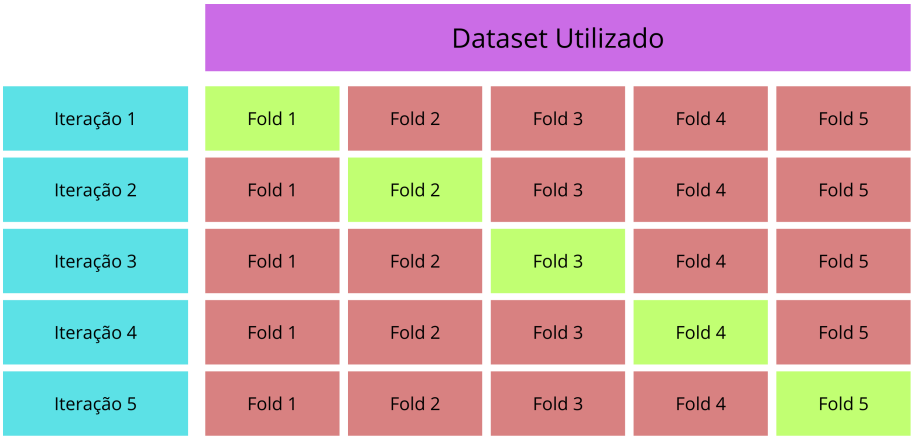
Uma vez completadas as K iterações, os resultados obtidos são combinados, geralmente calculando-se a média juntamente com o desvio padrão dos desempenhos obtidos, para produzir uma única medida de desempenho do modelo. Essa medida integrada fornece uma visão mais abrangente e confiável do desempenho do modelo, ao considerar sua performance em diferentes conjuntos de dados (CUNHA, 2019).

A Figura 2.9 ilustra um exemplo da utilização do *k-fold* com $k = 5$, ou seja, cinco partições.

2.1.8 *Métricas de Avaliação*

Para avaliar a performance de modelos de classificação, especialmente em problemas de *Machine Learning*, é essencial utilizar métricas de avaliação que proporcionem uma visão clara da qualidade das previsões do modelo. Entre as principais métricas utilizadas, destacam-se

Figura 2.9 – Demonstração de divisão e utilização dos *folds*, considerando $k = 5$, no qual a cada execução um fold será usado no teste.



Fonte: Próprio Autor.

a Matriz de Confusão, Acurácia, Precisão, Revocação e *F1-Score*.

Matriz de Confusão

A Matriz de Confusão é uma representação que permite visualizar o desempenho de um algoritmo de classificação. Ela compara as previsões do modelo com os valores reais, apresentando a quantidade de Verdadeiro Positivo (VP), Falso Positivo (FP), Verdadeiro Negativo (VN) e Falso Negativo (FN). A matriz de confusão é especialmente útil para identificar quais tipos de erros o modelo está cometendo, ajudando a direcionar ajustes e melhorias (MAGNO, 2023).

A forma mais simples de observar uma matriz de confusão é por meio de um problema binário, conforme mostrado na Tabela 2.1.

Tabela 2.1 – Exemplo de uma Matriz de Confusão.

Matriz	**Previsto Positivo**	**Previsto Negativo**
Real Positivo	VP	FN
Real Negativo	FP	VN

Fonte: Próprio autor.

Os elementos da Tabela 2.1 podem ser definidos da seguinte forma:

- **Verdadeiro Positivo (VP)**: Representa a quantidade de previsões corretas em que o modelo identificou corretamente a classe positiva.
- **Verdadeiro Negativo (VN)**: Mostra as previsões corretas em que o modelo identificou corretamente a classe negativa.
- **Falso Positivo (FP)**: Indica o número de previsões incorretas em que o modelo classificou como positiva uma classe que, na realidade, é negativa.

- **Falso Negativo (FN)**: Refere-se ao número de previsões incorretas em que o modelo não conseguiu identificar uma classe positiva, classificando-a como negativa.

Estes elementos também podem ser empregados para a definição das métricas de acurácia, precisão e revocação.

Acurácia

A Acurácia é uma métrica extremamente intuitiva e amplamente utilizada, pois oferece uma medida direta da proporção de previsões corretas em relação ao número total de casos avaliados. É uma métrica valiosa para entender a eficácia geral de um modelo de classificação. No entanto, é importante ressaltar que, em conjuntos de dados desbalanceados, onde uma classe é significativamente mais frequente do que a outra, a acurácia pode fornecer uma avaliação distorcida do desempenho do modelo. Isso ocorre porque a acurácia não leva em consideração a distribuição das classes no conjunto de dados, o que pode levar a uma interpretação equivocada da qualidade do modelo (BLOG, 2020). A fórmula para calcular a acurácia pode ser descrita por:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}. \tag{2.1}$$

Precisão

A Precisão, também conhecida como Valor Preditivo Positivo, é uma métrica crucial para avaliar a capacidade de um modelo de classificação em fazer previsões corretas entre os casos que ele previu como positivos. Em outras palavras, a precisão mede a proporção de verdadeiros positivos em relação ao número total de instâncias classificadas como positivas pelo modelo (BLOG, 2020).

É importante destacar que a precisão é especialmente relevante em situações em que os falsos positivos têm um custo significativo ou quando é fundamental garantir que as previsões positivas sejam verdadeiramente precisas. No entanto, a precisão não leva em consideração os verdadeiros positivos que foram erroneamente classificados como negativos (falsos negativos), o que pode ser um aspecto crítico em determinados contextos, como na área médica (BLOG, 2020).

A fórmula para calcular a precisão é dada por:

$$\text{Precisão} = \frac{VP}{VP + FP}. \tag{2.2}$$

Revocação

A Revocação, também conhecida como Sensibilidade, é uma métrica fundamental para avaliar a capacidade de um modelo de classificação em identificar corretamente todos os casos

positivos em um conjunto de dados. Em outras palavras, a revocação mede a proporção de verdadeiros positivos identificados pelo modelo em relação ao número total de casos positivos presentes no conjunto de dados (BLOG, 2020).

A revocação é especialmente importante em situações em que a identificação de todos os casos positivos é crucial e não pode haver falsos negativos. Por exemplo, na detecção de doenças graves, é essencial que o modelo identifique corretamente todos os casos positivos, mesmo que isso resulte em alguns falsos positivos (BLOG, 2020).

A fórmula para calcular a revocação é dada por:

$$\text{Revocação} = \frac{VP}{VP + FN}. \tag{2.3}$$

F1-Score

O *F1-Score* é uma métrica que combina a Precisão e a Revocação em uma única métrica, sendo especialmente útil em cenários onde existe um desequilíbrio entre as classes. Ele é a média harmônica entre a Precisão e a Revocação, garantindo que o modelo apresente um bom equilíbrio entre ambas. Isso significa que o *F1-Score* leva em consideração tanto os falsos positivos quanto os falsos negativos, oferecendo uma visão mais completa do desempenho do modelo de classificação (BLOG, 2020).

O *F1-Score* é particularmente relevante em situações em que tanto os falsos positivos quanto os falsos negativos têm custos elevados, como no diagnóstico médico, onde a precisão isolada pode não ser suficiente para avaliar o modelo de forma adequada. Através dessa métrica, é possível obter uma avaliação mais robusta, considerando os dois tipos de erro (BLOG, 2020).

A fórmula para calcular o F1-Score é dada por:

$$\text{F1-Score} = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}. \tag{2.4}$$

2.1.9 *Data Augmentation*

A técnica de *Data Augmentation* é essencial em *Machine Learning* especialmente em tarefas de visão computacional, que visa aumentar a diversidade e a quantidade de dados disponíveis para o treinamento de modelos. Essa abordagem é viável em situações em que os conjuntos de dados são limitados ou desbalanceados, o que pode comprometer a capacidade de generalização de um modelo. Ao criar variações artificiais dos dados originais, o *Data Augmentation* ajuda a prevenir problemas como *overfitting*, permitindo que os modelos aprendam de forma mais robusta (NALEPA; MARCINKIEWICZ; KAWULOK, 2019).

As transformações aplicadas para realizar o aumento de dados são normalmente projetadas para preservar a semântica dos dados, enquanto introduzem variações que simulam diferentes

condições do mundo real. É válido ressaltar que essas transformações podem incluir alterações geométricas, como rotações, translações, espelhamentos e redimensionamentos, bem como ajustes de cor, como mudanças na luminosidade, saturação e contraste. Além disso, em cenários onde as classes estão desbalanceadas, a aplicação dessas técnicas pode contribuir para um aprendizado mais equilibrado, mitigando vieses nos resultados (NALEPA; MARCINKIEWICZ; KAWULOK, 2019).

Apesar de seus benefícios, a técnica de *Data Augmentation* não é isenta de desafios. A escolha inadequada de transformações pode introduzir artefatos irreais ou distorções que comprometem a integridade dos dados, resultando em um treinamento menos eficaz. Além disso, a definição do conjunto ideal de transformações para um problema específico muitas vezes requer experimentação e ajustes, o que pode demandar tempo e recursos (NALEPA; MARCINKIEWICZ; KAWULOK, 2019).

Na Figura 2.10 é apresentada a imagem original e o resultado da aplicação da técnica de espelhamento horizontal em uma imagem cerebral.

Figura 2.10 – Amostra de imagem original e imagem espelhada em torno do eixo Y.

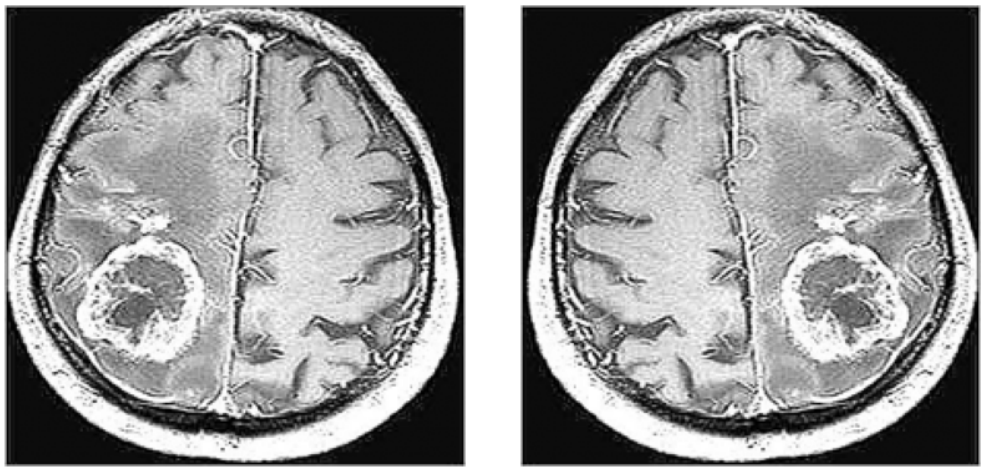


Imagem Normal Imagem Espelhada

Fonte: (ABHRANTA PANIGRAH, 2021).

2.1.10 *Cross-Dataset*

O *cross-dataset* é uma abordagem utilizada em *machine learning* para avaliar a capacidade de generalização de modelos. Essa estratégia consiste em treinar modelos em um conjunto de dados (chamados de *source dataset*) e testá-los em outro conjunto distinto (chamados de *target dataset*). A diferença entre os conjuntos de dados, geralmente presente em termos de distribuição, características das amostras, ou condições de captura, representa um desafio adicional para o desempenho dos modelos e reflete situações reais onde os dados de treinamento e teste não possuem a mesma origem (CHAO; HU; SHA, 2018).

O uso de *cross-dataset* é particularmente relevante em aplicações médicas, pois nesses casos, os dados podem variar significativamente devido a fatores como mudanças de equipamentos, condições ambientais e diversidade de populações. A capacidade de um modelo de manter um desempenho robusto em cenários tão diversos é importante para sua aplicabilidade no mundo real (RUNDO et al., 2020).

No entanto, o uso de *cross-dataset* também apresenta desafios. Diferenças substanciais entre os conjuntos podem levar à degradação do desempenho do modelo, um fenômeno conhecido como *dataset shift*. Esse problema ocorre porque o modelo, treinado em um conjunto de dados específico, tende a aprender padrões e distribuições particulares que podem não estar presentes no conjunto de teste. Como resultado, a capacidade de generalização do modelo é comprometida (RANFTL et al., 2020).

Para mitigar esses efeitos, diversas estratégias podem ser empregadas. Métodos de adaptação de domínio buscam alinhar as distribuições dos conjuntos de dados de origem e destino, enquanto técnicas de *transfer learning* permitem aproveitar o conhecimento adquirido no *source dataset* para melhorar o desempenho no *target dataset*. Além disso, o uso de normalizações apropriadas, seleção de características mais discriminativas e a inclusão de técnicas como *data augmentation* durante o treinamento podem ajudar a reduzir a disparidade entre os conjuntos de dados (RANFTL et al., 2020).

2.2 Trabalhos Relacionados

A pesquisa acerca da classificação de tumores cerebrais, mediante o emprego de técnicas de Deep Learning, está fundamentada em uma vasta gama de estudos, tais como Ullah et al. (2023) e Shabbir, Nazir et al. (2023), gerando progressos significativos na intersecção entre a medicina e a computação. A importância da precisão na detecção de patologias, especialmente em casos de tumores cerebrais, tem sido um ponto central em estudos que exploram a aplicação de técnicas de *Deep Learning* em radiologia. Tais investigações, como discutido de forma abrangente por Litjens et al. (2017) e Pereira et al. (2019), evidenciam o papel crucial dessas técnicas na melhoria do diagnóstico e tratamento de patologias cerebrais, fornecendo *insights* para os profissionais da saúde e os pesquisadores na área.

Litjens et al. (2017) oferecem uma visão abrangente dos conceitos fundamentais de *Deep Learning* no contexto da radiologia. Especificamente, o estudo destaca como as técnicas de *Deep Learning* podem ser aplicadas para aprimorar a detecção e classificação de condições médicas complexas, como tumores cerebrais, em imagens de RM e outros exames de imagem. As redes neurais profundas, utilizadas no contexto do *Deep Learning*, tem a capacidade de aprender padrões complexos e nuances nas imagens médicas, proporcionando uma vantagem significativa na detecção precoce e precisa de anomalias, como tumores cerebrais.

Além disso, a menção à revisão sistemática conduzida por Tajbakhsh et al. (2020) destaca a

continuidade do interesse e pesquisa na aplicação do *Deep Learning* em neuroimagem. As revisões sistemáticas têm um papel crucial ao consolidar os resultados de vários estudos, proporcionando uma visão geral do estado atual da pesquisa. É apresentado não apenas os avanços significativos na detecção e classificação de anomalias cerebrais por meio do *Deep Learning*, mas também oferecem uma análise crítica das metodologia empregadas nos estudos revisados.

Um estudo específico realizado por Havaei et al. (2017) aborda a aplicação de redes neurais profundas na classificação e segmentação de tumores cerebrais utilizando imagens de RM. Este trabalho apresenta uma arquitetura específica para esse proposito e destaca resultados promissores na precisão da classificação. A arquitetura desenvolvida pelos autores é projetada para otimizar a precisão na classificação de tumores cerebrais. A segmentação é uma tarefa crítica em imagens médicas, pois permite uma compreensão mais detalhada das regiões de interesse, contribuindo para diagnósticos mais precisos. Ao abordar a classificação e segmentação de tumores cerebrais, Havaei et al. (2017) contribuiu para o desenvolvimento de ferramentas mais avançadas e precisas no campo da medicina diagnóstica. O estudo não apenas destaca os resultados alcançados, mas também ressalta a importância do uso de redes neurais profundas, uma categoria de algoritmos de *Machine Learning*, na resolução de desafios complexos em imagens médicas.

Alrabai (2023) teve como objetivo do trabalho detectar tumores cerebrais usando CNN, com um modelo menos complexo composto por três camadas com algumas modificações como normalização, redimensionamento de imagens e aumento de dados na parte de pré-processamento. A partir dessas decisões, conseguiu-se um resultado de acurácia de 96% utilizando a base de dados nomeada como “Brain_Tumor_Detection_MRI”, descrita em (ABHRANTA PANIGRAH, 2021), demonstrando a importância desse trabalho.

Dentro do trabalho proposto por Lamrani et al. (2022), foi utilizado um modelo de CNN com quatro camadas convolucionais e três camadas de *maxpooling* alternadas, adicionadas ao final com seis camadas densas. Para evitar o *overfitting* do modelo foi-se utilizado tanto o aumento de dados quanto camadas de *dropout*, dentro das camadas ocultas do modelo. Dentro desse trabalho, utilizou-se quase 15 milhões de parâmetros. Com essa arquitetura, pode-se obter um resultado de acurácia de 96.33% na fase de teste utilizando a base de dados proposta em (ABHRANTA PANIGRAH, 2021).

No trabalho de (RAGHU et al., 2022), foi realizado uma análise comparativa entre CNN e ViT no que se refere à extração de características visuais. Para isso, foram conduzidos testes quantitativos e, principalmente, investigações sobre como as camadas internas de cada arquitetura trocam informações. Os experimentos mostraram que os ViTs não apenas produziram representações mais homogêneas entre suas camadas, mas também preservaram de maneira mais eficaz a informação espacial das imagens. A comparação quantitativa entre as duas arquiteturas foi realizada através do treinamento de diferentes modelos na base de dados JFT-300M (SUN et al., 2017), com a avaliação feita no conjunto de dados ImageNet (DENG et al., 2009), medindo a

acurácia de cada modelo. O ViT-H/14 (DOSOVITSKIY et al., 2020b) obteve um desempenho superior, com uma acurácia de 80%, aproximadamente 5% maior que a alcançada pela CNN ResNet152x2. Esses resultados demonstram uma performance superior dos modelos baseados em ViTs em cenários de avaliação *zero-shot* quando comparados às CNNs testadas.

O trabalho de Shabbir, Nazir et al. (2023) é baseado em duas arquiteturas de *Deep Learning*: *EfficientNet e ResNet50*. Foram empregadas técnicas de aumento de dados, como deformações elásticas e de escala, alteração de brilho, espelhamento e remoção de ruído, sendo esta última realizada com o uso do *Block-matching and 3D filtering* (BM3D) nas imagens dos tumores cerebrais. O objetivo era aumentar a capacidade do modelo de lidar com uma ampla gama de possíveis cenários clínicos e condições de imagem médica. Entre os resultados obtidos por Shabbir, Nazir et al. (2023), destaca-se a acurácia obtida de 96,73% utilizando a *EfficientNet*, enquanto a *ResNet50* alcançou 96,08% na fase de teste utilizando a base de dados de (ABHRANTA PANIGRAH, 2021).

Por outro lado, o estudo de (OQUAB et al., 2023b) explora o desempenho avançado dos ViT pré-treinados utilizando métodos de auto-supervisão na tarefa de classificação de imagens. Neste trabalho, os autores apresentam o DINOV2, um método de aprendizado auto-supervisionado aplicado a um conjunto diversificado de dados de treinamento. Para testar a eficácia na extração de características, os experimentos foram realizados com um modelo de classificação linear simples, utilizando um *backbone* de extração de características com os pesos congelados. Quando avaliado no conjunto de dados *ImageNet-1k* como *benchmark*, o modelo ViT-g/14 treinado com o DINOV2 alcançou uma acurácia de 86,5%, superando em 4,2% o antigo estado-da-arte, o iBOT ViT-L/16 (ZHOU et al., 2022).

Os resultados de (ZHOU et al., 2022) ressaltam o impacto do DINOV2 na melhoria do desempenho do *Visual Transformers* (ViT), abrindo portas para novas abordagens que busquem maximizar o potencial desses modelos. Dando continuidade a essa linha de pesquisa, (KIM et al., 2023) propõe aperfeiçoar os classificadores baseados em extratores de características no paradigma *zero-shot*, com foco nas capacidades espaciais dos ViTs. Neste trabalho, o *backbone* do extrator ViT permanece congelado, mas são incluídas camadas adicionais lineares e de atenção que realizam um tipo de pós-processamento das características extraídas. O objetivo é captar não apenas os atributos visuais, mas também as relações entre esses atributos. Após o pós-processamento, um classificador linear é treinado com as saídas geradas. Testado em diversos *benchmarks*, esse método alcançou uma acurácia harmônica de 67,9% na base de dados CUB (HE; PENG, 2020), resultando em uma melhoria de cerca de 13% em relação ao estado-da-arte com CNNs, além de um ganho mínimo de 8% em comparação com outros métodos baseados em paradigmas *zero-shot* generalizados.

Ademais, o trabalho realizado por Ullah et al. (2023) destaca avanços significativos na detecção de tumores cerebrais, apresentando a arquitetura TumorDetNet como uma contribuição proeminente. A precisão alcançada de 99.83% no conjunto de dados disponível em (ABHRANTA

PANIGRAH, 2021), reflete a eficácia dessa abordagem na identificação precisa de formações tumorais em imagens médicas. A arquitetura TumorDetNet, caracterizada por suas 48 camadas de ativação RELU, surge como um componente crucial para o sucesso alcançado no processo de detecção. A escolha e otimização dessas camadas demonstram um profundo entendimento das nuances e complexidades presentes nas imagens de RM, evidenciando a sofisticação do modelo proposto.

O presente trabalho possui semelhanças com os anteriores, alinhando-se a conceitos de *Deep Learning*. Entretanto, destaca-se neste estudo a utilização de abordagens *zero-shot* ao invés da utilização de abordagens tradicionais como CNN. Estas abordagens permitem que o modelo seja aplicado diretamente em novas tarefas de detecção de tumores sem a necessidade de treinamento de todo um modelo. A vantagem de uma abordagem *zero-shot* está na sua capacidade de extrair características descritivas sem a necessidade de um *fine-tunning*, tornando-o uma possível solução para a detecção de tumores em diversos contextos clínicos.

3 Base de Dados e Abordagem Proposta

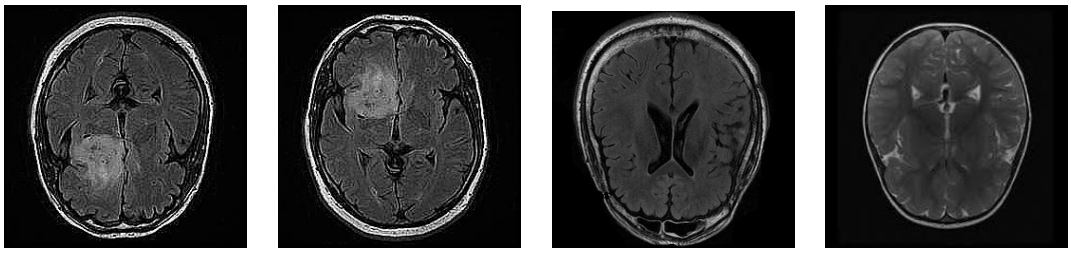
Neste capítulo, serão apresentadas as bases de dados que foram utilizada na Seção 3.1 e a metodologia feita na Seção 3.2.

3.1 Base de Dados

Este trabalho utiliza, primeiramente, a base de dados Brain Tumor Detection MRI (BTDM) (ABHRANTA PANIGRAH, 2021), disponível no site Kaggle[1], que abrange um compilado de imagens cerebrais capturadas por meio de RM, com o intuito de identificar a existência ou não de tumores cerebrais. O conjunto de dados é composto por um total de 3.000 imagens correspondentes, no qual metade dos dados (1.500 imagens) são de cérebros que contêm tumores e a outra metade (1.500 imagens), representa cérebros que não contêm tumores.

Na Figura 3.1 abaixo, pode-se observar alguns exemplos de imagens presentes no conjunto de dados.

Figura 3.1 – Amostras de imagens da base de dados. As imagens (a) e (b) representam imagens de tumor cerebral, enquanto as imagens (c) e (d), sem tumor.



(a) (b) (c) (d)

Fonte: (ABHRANTA PANIGRAH, 2021).

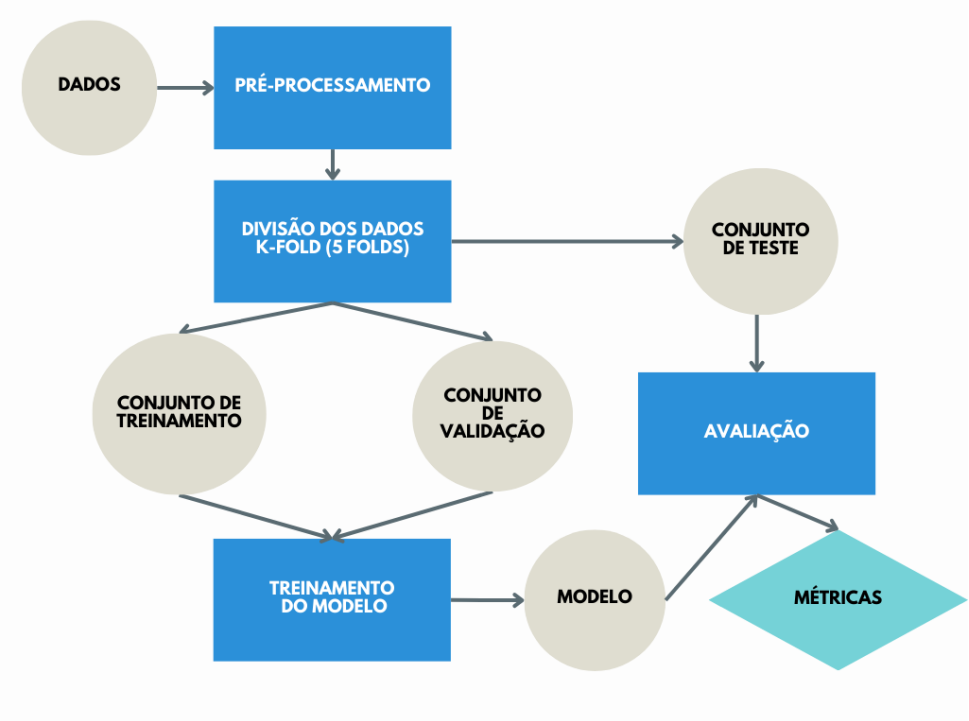
Além da base de dados supracitada, a base de dados Brain MRI Images for Brain Tumor Detection (BMI) (Navoneel Chakrabarty, 2019), disponível no site Kaggle[2], também foi utilizada. Diferente da base de dados anterior, esta base abrange uma quantidade menor de imagens cerebrais capturadas por meio de RM. Tal base de dados é composto por um total de 253 imagens correspondentes, no qual 155 imagens são de cérebros que contêm tumores enquanto 98 imagens representam cérebros que não contêm tumores.

[1] Disponível em: <https://www.kaggle.com/datasets/abhranta/brain-tumor-detection-mri>.
[2] Disponível em: <https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection>.

3.2 Metodologia

Esta seção demonstra a metodologia utilizada nesse trabalho. Ela pode ser dividida em quatro etapas principais: pré-processamento dos dados, divisão da base de dados, treinamento do modelo e avaliação do modelo. Na Figura 3.2, é representado uma visão sobre as etapas do modelo adotado nesse trabalho. Inicialmente, os dados são obtidos da base de dados "*Brain Tumor Detection MRI*" (ABHRANTA PANIGRAH, 2021). A partir desses dados, as imagens passam por um pré-processamento, onde são aplicados os métodos de normalização, *data augmentation* e redimensionamento, para prepará-las de forma adequada para o reconhecimento dos tumores cerebrais. Em seguida, as imagens são divididas em três categorias: treinamento, validação e teste. Dentro dessa divisão, é aplicado a estratégia de avaliação *K-Fold*, no qual, a cada iteração, é aplicado um *fold* diferente no conjunto de teste. O modelo proposto é então treinado usando os conjuntos de treinamento e validação. Após o treinamento, o modelo, agora adaptado, é utilizado para classificar as imagens do conjunto de teste. Por fim, o modelo realiza suas previsões, fornecendo os resultados de classificação das imagens avaliadas.

Figura 3.2 – Fluxograma do modelo, no qual é descrito as etapas a cada evolução do modelo.



Fonte: Próprio Autor.

3.2.1 Pré-Processamento de Dados

No primeiro passo, as imagens retiradas das bases de dados passam por um processo de *data augmentation*, utilizando espelhamento lateral. Em seguida, é aplicado um processo de normalização, no qual os valores dos *pixels* das imagens são ajustados para um intervalo específico entre 0 e 1, garantindo que todas as imagens de entrada estejam na mesma escala. A

normalização é feita dividindo-se cada *pixel* pelo pixel de maior intensidade, o que assegura que cada valor de *pixel* esteja entre 0 e 1.

Por último, as imagens são redimensionadas para uma dimensão específica de (224×224). Esse processo utiliza o método de interpolação por área, que ajusta o tamanho da imagem preservando detalhes visuais importantes através de uma média ponderada. Esse redimensionamento padroniza as imagens para terem tamanho exigido pelos modelo treinados neste trabalho.

Após isso, os dados e suas respectivas classes foram embaralhados aleatoriamente antes de serem divididos em conjuntos de treinamento, validação e teste.

3.2.2 Divisão da Base de Dados

A partir desse conjunto de dados, foi utilizado a estratégia de *K-Fold*, mais precisamente do *StratifiedKFold* (MAYANGSARI; SYARIF; BARAKBAH, 2023), com $k = 5$ *folds*, no qual divide-se o conjunto de dados com cada hora um *fold* sendo parte do conjunto de teste, enquanto os demais são do conjunto de treinamento, de forma a ter a mesma distribuição de classes. Dentro desse modelo, utilizou-se 72% dos dados para treinamento, 8% para validação e 20% para teste. A Tabela 3.1 mostra a distribuição de cada classe dentro de cada conjunto do *K-Fold* para ambas as bases de dados usadas neste trabalho.

Tabela 3.1 – Distribuição de imagens por classe nos conjuntos de treinamento, validação e teste em cada iteração do *K-Fold* das bases de dados "*Brain Tumor Detection MRI*" e "*Brain MRI Images for Brain Tumor Detection*".

Base de Dados	Classe	Treinamento	Validação	Teste
Brain Tumor Detection MRI - BTDM	"Sim"	1080	120	300
	"Não"	1080	120	300
	Total	2160	240	600
Brain MRI Images for Brain Tumor Detection - BMI	"Sim"	112	12	31
	"Não"	70	8	20
	Total	182	20	31

Fonte: Próprio autor.

3.2.3 Treinamento do Modelo

Essa seção tem como objetivo avaliar duas estratégias diferentes, no qual uma é utilizando modelos baseados em CNNs *fine-tuned* e a outra baseada em modelos com base na extração de características seguindo a estratégia *zero-shot*.

3.2.3.1 Treinamento de modelo *fine-tuning*

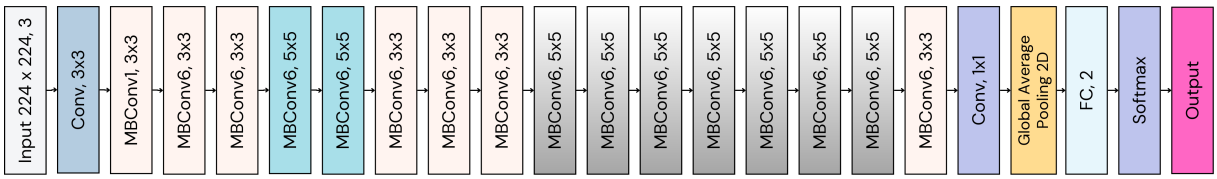
Na busca por melhorar o desempenho na tarefa de classificação de imagens, optou-se por utilizar CNNs convencionais aplicando a técnica de *fine-tuning*. Diferentemente de abordagens que visam economizar recursos computacionais, o foco dessa estratégia é refinar os pesos do

modelo para minimizar o seu erro. Essa técnica permitiu ajustar a rede a partir de uma base pré-estabelecida, aprimorando o aprendizado de padrões específicos do novo conjunto de dados e melhorando a capacidade do modelo em capturar detalhes relevantes para a tarefa.

O *fine-tuning* do modelo foi aplicado com o intuito de extrair mais potencial da arquitetura escolhida, explorando sua capacidade de aprendizado em um contexto mais específico, sem recorrer a um modelo treinado do zero. Dessa forma, foi possível adaptar o comportamento da rede e tentar superar eventuais limitações de performance inerentes ao treinamento padrão.

Dentro dos modelos de CNNs, as arquiteturas escolhidas foram a *EfficientNet*-B0 (TAN; LE, 2019), a qual contém a estrutura base que pode ser vista na Figura 3.3, e a *EfficientNet*-V2 L. Elas oferecem um consumo baixo de recursos computacionais enquanto mantêm o desempenho preditivo em tarefas de visão computacional (SHABBIR; NAZIR et al., 2023).

Figura 3.3 – Arquitetura base da *EfficientNet*-B0.



Fonte: Próprio Autor

Tendo por base essas arquiteturas, adicionou-se duas camadas densas com ativação *ReLU* de 256 e 128 neurônios em ambas arquiteturas, respectivamente. Após isso, adicionou-se uma camada de *dropout* com 30% dos neurônios, e, por fim, mais duas camadas densas, ambas com ativação *ReLU* e 32 neurônios. A última camada do modelo é uma camada densa com dois neurônios e ativação *softmax*. Essa proposta de camadas adicionais foi baseada no trabalho previamente realizado em (JúNIOR; PAES; SILVA, 2024).

3.2.3.2 Treinamento de abordagem *Zero-Shot*

Em contraposto à estratégia de *fine-tuning*, optou-se também por explorar a técnica de extração de características e usá-las no paradigma *zero-shot*. Diferentemente do ajuste completo dos pesos da rede, essa abordagem visa utilizar modelos previamente treinados em grandes bases de dados, como a *ImageNet*, sem a necessidade de novos ciclos extensivos de treinamento. O foco, neste caso, foi reaproveitar diretamente os extratores de características, adicionando apenas uma camada de classificação, o que permite acelerar o processo de treinamento e economizar recursos computacionais.

Ao contrário da estratégia anterior, que buscava minimizar o erro do modelo refinando o aprendizado do modelo, a extração de características com o paradigma *zero-shot* foi empregada para avaliar o desempenho de arquiteturas que já haviam sido otimizadas para outras tarefas. Assim, foi possível comparar o uso de CNNs e ViTs como extratores. Desta forma, investigou-se

até que ponto os modelos podem generalizar para o novo problema de classificação sem qualquer ajuste adicional nos seus pesos, além de uma camada classificadora.

Para avaliar a capacidade de extratores de características utilizando o paradigma *zero-shot*, foram utilizados os seguintes modelos: **VT-FPN** (ViT treinado na *ImageNet-21*) (WU et al., 2020); o **DINOv2** (OQUAB et al., 2023b); a ***EfficientNet*-B0** (TAN; LE, 2019); e a ***EfficientNet*-V2 L**.

3.2.4 Avaliação Do Modelo

Depois de treinado, utilizou-se o modelo pra trabalhar na precisão dos dados de teste em cada *fold*, permitindo a avaliação da sua capacidade de generalização. A acurácia, precisão, revocação e F1-*Score* foram registradas para cada iteração, seguindo as Equações (3.1), (3.2), (3.3) e (3.4) abaixo, a partir da comparação das predições com as reais classes do modelo. Ao final, a média e o desvio padrão das acurácias foram calculadas para fornecer uma visão geral da evolução e desempenho do modelo a cada iteração.

$$\text{Média da Acurácia} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{\text{Número de predições corretas no } \textit{fold } i}{\text{Número total de exemplos no } \textit{fold } i}\right). \tag{3.1}$$

$$\text{Média da Precisão} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{\text{Verdadeiros Positivos no } \textit{fold } i}{\text{Verdadeiros Positivos no } \textit{fold } i + \text{Falsos Positivos no } \textit{fold } i}\right). \tag{3.2}$$

$$\text{Média da Revocação} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{\text{Verdadeiros Positivos no } \textit{fold } i}{\text{Verdadeiros Positivos no } \textit{fold } i + \text{Falsos Negativos no } \textit{fold } i}\right). \tag{3.3}$$

$$\text{Média do F1-Score} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{2 \times \text{Precisão no } \textit{fold } i \times \text{Revocação no } \textit{fold } i}{\text{Precisão no } \textit{fold } i + \text{Revocação no } \textit{fold } i}\right). \tag{3.4}$$

3.2.4.1 Avaliação *cross-dataset*

A avaliação *cross-dataset* foi conduzida com o objetivo de examinar a capacidade de generalização do modelo para bases de dados distintas, garantindo que o desempenho obtido não seja restrito a um único conjunto de dados. Essa abordagem é particularmente relevante para cenários onde a distribuição dos dados de treinamento pode diferir significativamente daquela encontrada em novos conjuntos de teste, como ocorre frequentemente em aplicações médicas e biomédicas.

Para essa avaliação, foram considerados os dois conjuntos de dados distintos: BTDM e BMI. Os modelos foram treinados separadamente em cada uma dessas bases, utilizando a estratégia de *K-Fold*, e testados utilizando toda a outra base de dados, ou seja:

- Modelos treinados na base BTDM foram avaliados na base BMI completa.
- Modelos treinados na base BMI foram avaliados na base BTDM completa.

Os resultados foram analisados utilizando as métricas de desempenho padrão: acurácia, precisão, revocação e *F1-score*. Ademais, foram reportados média e desvio para cada conjunto de experimentos.

4 Experimentos e Resultados

Esta seção apresenta os resultados obtidos com as estratégias descritas no Capítulo 3 e utilizando os modelos baseados em CNNs e extratores *Zero-Shot*.

Para a execução dos algoritmos, utilizou-se uma máquina equipada com 128Gb de memória RAM DDR4, GPU RTX 3090 e o processador Intel i9-10900. A implementação foi realizada utilizando a linguagem Python em conjunto com o *framework* TensorFlow para o modelo da *EfficientNet*-B0 e EfficientNet-V2 L e o *framework* PyTorch para o modelo VT-FPN e DINOV2. Para o treinamento do modelo, utilizou-se 15 épocas com *learning rate* inicial de 10^{-3} e decaimento exponencial a uma taxa de 0,9. Utilizou-se a função de perda *Sparse Categorical Crossentropy* e o protocolo de redução de *learning rate* seguindo a estratégia de *Exponential Decay*. O otimizador dos pesos usado foi o Adam. Por fim, a biblioteca Scikit-Learn foi utilizada para calcular as métricas dos modelos avaliados.

O presente estudo avaliou algumas arquiteturas utilizando tanto o paradigma *zero-shot*, quanto realizando um *fine-tunning* do modelo. Para critério de referência, o trabalho de Ullah et al. (2023) (arquitetura *TumorDetNet*) foi reimplementado para comparação tanto utilizando o protocolo proposto pelos autores, quanto o utilizado neste trabalho (*k-fold*, com k=5).

4.1 Resultados Obtidos

Na Tabela 4.1 são apresentados os resultados dos modelos baseados em CNNs convencionais tanto no paradigma de extrator *zero-shot* quanto no *fine-tuned* utilizando a base de dados BTDM. Avaliou-se modelos baseado em ViT no paradigma de *zero-shot*. Além disso, também foi implementado o modelo da *TumorDetNet* (ULLAH et al., 2023), tanto com a utilização de `train_test_split`, quanto com a do *K-Fold*. Serão apresentados modelos com *fine-tuned* como modelos "**FT**" e modelos com extração *Zero-Shot* como modelos "**ZS**" (essa nomenclatura será usada no restante do texto).

Como pode ser visto na Tabela 4.1, a proposta do ViT como extrator de características, tanto para o modelo VT-FPN quanto para o modelo DINOV2 apresentou desempenho melhor em relação aos modelo convencionais de CNN com *fine-tuning*. Avaliando os resultados obtidos, nota-se uma melhora do modelo VT-FPN e do modelo DINOV2 de 0,8% a 0,9% em relação ao modelo com *fine-tuning* da *EfficientNet*-B0. Além do ganho evidente de desempenho em todas as métricas, nota-se uma redução do desvio padrão em torno de 1,30% a 1,20%, evidenciando uma maior constância das métricas utilizadas.

Além do ganho evidente de desempenho expresso pelas métricas obtidas, também pode-se ressaltar que o processo de adequação dos pesos da camada de classificação das propostas

Tabela 4.1 – Resultados da média de Acurácia, Revocação, Precisão e F1-Score e o desvio padrão para cada uma das métricas na base BTDM. Melhores resultados realçados em negrito.

Modelo	**Acurácia (%)**	**Revocação (%)**	**Precisão (%)**	**F1-Score (%)**
TumorDetNet - `train_test_split`	**99,83**	**99,66**	**100,00**	**99,83**
TumorDetNet - K-Fold	96,50 ± 2,51	96,97 ± 2,70	96,49 ± 2,84	99,83 ± 3,01
FT *EfficientNet*-B0	97,60 ± 1,60	97,61 ± 1,60	97,60 ± 1,60	97,59 ± 1,60
FT *EfficientNet*-V2 L	50,23 ± 0,08	23,23 ± 0,08	50,23 ± 0,08	33,59 ± 0,09
FT DINOv2	98.43 ± 0.62	98.43 ± 0.61	98.43 ± 0.61	98.43 ± 0.62
ZS *EfficientNet*-B0	54,23 ± 7,96	34,33 ± 18,17	54,23 ± 7,96	40,88 ± 14,55
ZS *EfficientNet*-V2 L	75,36 ± 7,45	80,88 ± 1,95	75,36 ± 7,45	73,58 ± 10,17
ZS VT-FPN	98,43 ± 0,30	98,44 ± 0,30	98,43 ± 0,30	98,43 ± 0,30
ZS DINOv2	98,46 ± 0,38	98,47 ± 0,38	98,46 ± 0,38	98,46 ± 0,38
ZS DINOv2 com *Data Augmentation*	**99,15 ± 0,19**	**99,15 ± 0,19**	**99,14 ± 0,19**	**99,14 ± 0,19**

Fonte: Próprio autor.

zero-shot foram consideravelmente menos custosas computacionalmente que o processo de treinamento da *EfficientNet*-B0 *fine-tuned* e da *EfficientNet*-V2 L. Isso ocorre pois o bloco de extração de características não requer atualização de pesos, sendo necessária apenas uma rápida adequação na camada classificadora.

Ao analisar as propostas de extração de características, nota-se que o modelo baseado em ViT tanto do VT-FPN quanto do DINOV2 obteve um desempenho superior à *EfficienteNet*-B0 no contexto de *zero-shot*. Essa diferença se baseia no ganho significante nas métricas, o qual teve uma variação entre 40% e 50% como também no desvio padrão, que teve uma variação de 7,50% a 18,0%.

Além disso, foi aplicada a técnica de aumento de dados no modelo DINOV2, que obteve o melhor desempenho entre os modelos *Zero-Shot*, com o objetivo de avaliar a possibilidade de aprimorar os resultados. Com isso, conseguiu-se uma melhoria de 0,69% no desempenho do modelo com *Data Augmentation* em comparação à base de dados original, o qual demonstra a eficácia da abordagem adotada.

Os resultados superiores dos modelos VT-FPN e DINOV2 em comparação às CNNs convencionais podem ser atribuídos à sua capacidade de generalização e eficiência no processamento dos dados. Esses modelos apresentaram métricas mais elevadas e consistentes, o que evidencia sua robustez em tarefas complexas como classificação de imagens médicas. Além disso, o paradigma *zero-shot* na extração de características contribuiu para reduzir o custo computacional, mostrando-se uma abordagem vantajosa em cenários com dados limitados ou restrições de recursos. A menor variabilidade dos resultados, indicada pelo desvio padrão reduzido, reforça a confiabilidade dos modelos, o que é essencial em aplicações críticas.

A Tabela 4.2 apresentada destaca as diferenças no tempo de treinamento e inferência entre os modelos VT-FPN e DINOV2. O modelo DINOV2 demonstrou maior eficiência no treinamento, com uma redução de aproximadamente 13% no tempo em relação ao VT-FPN. Os tempos foram calculados com base no tempo total de execução para o treinamento de cada

Tabela 4.2 – Resultados do tempo de treinamento dos modelos e de inferência das imagens individuais.

Modelo	**Tempo de Treinamento**	**Tempo Inferência em CPU (Por imagem)**
TumorDetNet	1533 minutos	1,70 segundos
FT *EfficientNet*-B0	110 minutos	1,21 segundos
ZS VT-FPN	61 segundos	**0,91 segundos**
ZS DINOv2	**53 segundos**	0,95 segundos

modelo, considerando o uso exclusivo de CPU. Em contrapartida, o VT-FPN apresentou um tempo de inferência ligeiramente mais rápido por imagem.

A partir desses resultados, foi selecionado os seguintes modelos para serem testados na base de dados BMI e no teste de *Cross-Dataset*:

- ***TumorDetNet*** - `train_test_split` e ***TumorDetNet*** - *K-Fold*, representando o estado-da-arte.
- **FT *EfficientNet*-B0** e **FT DINOv2**, por serem os melhores modelos *fine-tuned*.
- **ZS DINOv2** e **ZS DINOv2 *Data Augmentation***, por serem os melhores modelos *Zero-Shot*.

Na Tabela 4.3 são apresentados os resultados dos modelos selecionados utilizando a base de dados BMI.

Tabela 4.3 – Resultados da média de Acurácia, Revocação, Precisão e F1-Score e o desvio padrão para cada uma das métricas na base BMI. Melhores resultados realçados em negrito.

Modelo	**Acurácia (%)**	**Revocação (%)**	**Precisão (%)**	**F1-Score (%)**
TumorDetNet - `train_test_split`	**96,08**	**95,00**	**95,00**	**95,00**
TumorDetNet - K-Fold	94,32 ± 1,13	95,12 ± 1,32	94,32 ± 1,13	93,89 ± 1,22
FT *EfficientNet*-B0	46,85 ± 11,47	23,27 ± 11,29	46,85 ± 11,47	30,71 ± 12,57
FT DINOv2	70,99 ± 11,51	66,23 ± 19,55	70,99 ± 11,51	66,67 ± 16,20
ZS DINOv2	86,06 ± 5,48	87,10 ± 6,07	86,06 ± 5,48	86,01 ± 5,65
ZS DINOv2 com *Data Augmentation*	84,42 ± 2,12	85,88 ± 2,03	84,42 ± 2,12	84,22 ± 2,36

Fonte: Próprio autor.

Como pode ser visto na Tabela 4.3, a proposta de extratores de características também obteve melhores resultados em relação aos modelos *fine-tuned*. Esses resultados elucidam uma melhoria do modelo DINOV2 em relação ao modelo da *EfficientNet*-B0 *fine-tuned* quanto a do DINOV2 *fine-tuned*, tendo uma melhoria de 39,21% e 15,07%, respectivamente. Além do ganho evidente de desempenho em todas as métricas, nota-se uma redução do desvio padrão em torno de 5,99% e 6,03%, respectivamente.

Além disso, foi utilizado o modelo DINOV2 com a técnica de aumento de dados, que, embora obteve um resultado de 1,64%, teve uma maior constância dos valores obtidos, por oferecer um desvio padrão de 3,36% menor.

4.1.1 Análise *Cross-dataset*

Como apresentado na Tabela 4.4, a abordagem que utiliza extratores de características continua a superar os modelos *fine-tuned* com os modelos sendo treinados na base de dados BMI e testados na BTDM. A comparação entre o modelo DINOV2 e o modelo *fine-tuned* da *EfficientNet*-B0 revela uma melhoria significativa de 34,93% em acurácia, em comparação com um aumento de 34,88% de acurácia em comparação ao DINOV2 *fine-tuned*.

Tabela 4.4 – Resultados da média de Acurácia, Revocação, Precisão e F1-Score e o desvio padrão para cada uma das métricas treinada na base BMI e testados na BTDM. Melhores resultados realçados em negrito.

Modelo	**Acurácia (%)**	**Revocação (%)**	**Precisão (%)**	**F1-Score (%)**
TumorDetNet - `train_test_split`	83,19	86,31	83,19	84,83
TumorDetNet - K-Fold	73,12 ± 1,53	75,19 ± 2,12	73,12 ± 1,53	74,15 ± 1,56
FT *EfficientNet*-B0	49,95 ± 0,22	24,95 ± 0,22	49,95 ± 0,22	33,28 ± 0,25
FT DINOv2	50,00 ± 0,00	25,00 ± 0,00	50,00 ± 0,00	33,28 ± 0,00
ZS DINOv2	84,88 ± 2,33	87,31 ± 1,03	84,88 ± 2,33	84,59 ± 2,52
ZS DINOv2 *Data Augmentation*	**89,19 ± 2,68**	**90,04 ± 1,69**	**89,20 ± 2,68**	**89,11 ± 2,79**

Fonte: Próprio autor.

Ademais, o modelo DINOV2 foi novamente testado com a técnica de aumento de dados, o que resultou em um ganho adicional de 4,31%. Embora esse aumento não tenha sido tão expressivo quanto os resultados obtidos nos modelos principais, a aplicação de aumento de dados ainda demonstrou uma melhoria no desempenho e uma redução na variação dos resultados, sugerindo que o modelo com essa técnica apresentou maior consistência nas métricas avaliadas.

Por fim, na Tabela 4.5 são apresentados os resultados dos modelos treinados na base de dados BTDM e testados na base de dados BMI.

Tabela 4.5 – Resultados da média de Acurácia, Revocação, Precisão e F1-Score e o desvio padrão para cada uma das métricas treinada na base BTDM e testados na BMI. Melhores resultados realçados em negrito.

Modelo	**Acurácia (%)**	**Revocação (%)**	**Precisão (%)**	**F1-Score (%)**
TumorDetNet - `train_test_split`	86,73	88,12	86,73	87,19
TumorDetNet - K-Fold	84,53 ± 1,53	87,03 ± 2,31	85,32 ± 1,72	84,53 ± 1,53
FT *EfficientNet*-B0	52,83 ± 17,68	46,59 ± 31,06	52,83 ± 17,68	40,71 ± 23,63
FT DINOv2	49,53 ± 0,39	100,00 ± 0,00	49,53 ± 0,39	66,25 ± 0,35
ZS DINOv2	87,63 ± 1,07	90,18 ± 1,35	87,63 ± 1,07	87,79 ± 1,06
ZS DINOv2 *Data Augmentation*	**87,86 ± 0,96**	**90,80 ± 0,54**	**87,86 ± 0,96**	**88,03 ± 0,95**

Fonte: Próprio autor.

A Tabela 4.5 evidencia que a abordagem baseada em extração de características obteve um desempenho superior aos modelos *fine-tuned*. O modelo DINOV2 na configuração *zero-shot* apresentou uma acurácia de 87,63%, contrastando com os 52,83% alcançados pela *EfficientNet*-B0 *fine-tuned* e os 49,53% do DINOV2 *fine-tuned*, correspondendo a uma diferença de 34,80% e 38,10%, respectivamente. Apesar do baixo desvio padrão na acurácia (0,39%), o DINOV2

fine-tuned demonstrou um comportamento enviesado, com revocação de 100%, sugerindo uma tendência a classificar todas as amostras como pertencentes à classe predominante.

Ao empregar *data augmentation*, o modelo DINOV2 obteve um leve incremento na acurácia, atingindo 87,86%. Mais relevante, no entanto, foi a redução na variabilidade dos resultados, com o desvio padrão da revocação diminuindo de 1,35% para 0,54%. Esse efeito indica que a aplicação de técnicas de aumento de dados em bases pequenas não apenas manteve o alto desempenho do modelo, como também aumentou sua estabilidade, reduzindo flutuações entre as execuções.

4.2 Comparação com Estado da Arte

O modelo *TumorDetNet* (ULLAH et al., 2023) representa o estado da arte para as bases de dados utilizadas neste estudo, alcançando uma acurácia de 99,83% na base BTDM e de 96,08% na base BMI, o que supera em 0,69% e 10,02%, respectivamente, dos melhores resultados obtidos neste trabalho. No entanto, embora tenha sido aplicada a mesma proporção de divisão de dados para treinamento (80%) e teste (20%) não foi empregada uma estratégia de validação cruzada. Além disso, o modelo *TumorDetNet* apresenta uma arquitetura mais complexa em termos de número de camadas e parâmetros em relação ao trabalho proposto, o que pode resultar em maior demanda computacional e tempo de treinamento.

Para fins de comparação, os autores de Ullah et al. (2023) relataram um tempo de 1.533 minutos para o treinamento em uma CPU (Intel(R) Core(TM) i5-5200U), enquanto o modelo de classificação baseados nas características extraídas com o DINOV2 foram treinados em apenas 53 segundos. Isso demonstra o baixo custo computacional necessário para treinar, adaptar ou criar novos modelos. No que se refere ao tempo de inferência também em CPU, o DINOV2 leva em média 0,95 segundos para processar uma imagem (tempo médio para 100 imagens), mantendo-se viável para o uso diário em um hospital. Em contrapartida, o modelo *TumorDetNet* (ULLAH et al., 2023) apresenta um tempo médio de 1,7 segundos por imagem no mesmo *setup*, quase duas vezes mais lento que a abordagem proposta neste estudo, sem oferecer um ganho significativo em desempenho para o mesmo conjunto no mesmo *hardware*.

É importante destacar que a abordagem proposta neste trabalho sobressai pelo baixo custo computacional necessário para o treinamento, além de um tempo de inferência viável em contextos onde o processamento na ordem de segundos é aceitável. Como o extrator ViT não é treinado, a atualização dos pesos do modelo de classificação é realizada com maior frequência e sem a necessidade de recursos computacionais significativos. Isso facilita a disseminação desses modelos em diversas aplicações na área da saúde, devido ao baixo custo associado ao treinamento.

Nos experimentos de *cross-dataset*, os modelos *zero-shot* baseados no DINOV2 superaram o *TumorDetNet* em ambas as direções de avaliação. No teste realizado com modelos treinados na base BMI e avaliados na BTDM, o DINOV2 com *data augmentation* obteve a melhor acurácia

(89,19%) e *F1-Score* (89,11%), superando o *TumorDetNet*, cuja melhor configuração atingiu apenas 83,19% e 84,83%, respectivamente. Além disso, a revocação do DINOV2 com *data augmentation* (90,04%) foi superior à do *TumorDetNet* (86,31%), indicando uma maior capacidade de identificar corretamente os casos positivos.

No cenário inverso, com modelos treinados na BTDM e testados na BMI, o DINOV2 novamente apresentou melhor desempenho, atingindo 87,86% de acurácia e 88,03% de *F1-Score*, enquanto o *TumorDetNet* obteve 86,73% e 87,19%. A revocação do modelo *zero-shot* (90,80%) também superou a do *TumorDetNet* (88,12%), consolidando sua vantagem na detecção de tumores.

Esses resultados demonstram que os modelos que utilizam o paradigma de extração *zero-shot* não apenas atingem um desempenho superior ao do *TumorDetNet* nos dois cenários de avaliação, mas também possuem uma capacidade de generalização mais robusta entre diferentes conjuntos de dados, sem a necessidade de re-treinamento extensivo. Isso evidencia a eficiência dos métodos extratores no paradigma *zero-shot*, tornando-os alternativas mais viáveis para aplicações em contextos clínicos, onde a adaptação a novos domínios de dados é essencial.

5 Considerações Finais

Este capítulo apresenta as conclusões alcançadas ao longo do estudo, destacando os objetivos cumpridos e os resultados obtidos nos experimentos realizados. Além disso, são sugeridas direções para a continuidade desta pesquisa, considerando as descobertas e os resultados alcançados ao longo do trabalho.

5.1 Conclusão

Este trabalho teve como objetivo avaliar modelos baseados no paradigma *zero-shot* em comparação com CNNs convencionais com *fine-tuning*, a fim de verificar se é possível alcançar resultados satisfatórios na detecção de tumores cerebrais com baixo custo computacional. Os resultados revelaram que os modelos *zero-shot* VT-FPN e DINOV2 não apenas superaram a *EfficientNet*-B0 e a *EfficientNet*-V2 L em termos de acurácia, precisão, revocação e F1-Score, como também apresentou menor variabilidade nos resultados, com um desvio padrão mais baixo.

Os extratores de características demonstraram um desempenho positivo em relação ao modelo *EfficientNet*-B0 com *fine-tuning*, com o modelo DINOV2 alcançando uma acurácia de 98,46% e 99,15% com o uso de *data augmentation*. Além disso, o desvio padrão da *EfficientNet*-B0 foi superior ao do VT-FPN, com diferença de 1,4%

Além do desempenho superior, os modelos dos extratores no paradigma *zero-shot* proporcionaram uma significativa redução no custo computacional no tempo de treinamento. Essa economia é crucial para aplicações hospitalares, onde tempo e recursos são frequentemente limitados. A comparação com o estado da arte, representado pelo *TumorDetNet* (ULLAH et al., 2023), mostrou que, embora este último tenha apresentado um pequeno ganho em acurácia, a abordagem proposta neste estudo se destaca por sua rapidez e menor demanda computacional, mantendo a viabilidade para uso clínico.

5.2 Trabalhos Futuros

Como trabalhos futuros, sugere-se a avaliação da incerteza dos modelos treinados utilizando técnicas baseadas em *Conformal Prediction*. Essa abordagem permitiria quantificar a incerteza associada às classificações, fornecendo intervalos de confiança para cada predição e tornando o modelo mais interpretável em cenários clínicos. A incorporação dessa técnica pode aprimorar a robustez das decisões automatizadas, auxiliando profissionais da saúde na identificação de casos onde a incerteza do modelo exige uma análise mais detalhada.

Ademais, o uso de técnicas de interpretabilidade dos modelos treinados é outra aná-

lise interessante. Uma vez que elas podem ser mais um auxílio no diagnóstico de médicos e especialistas, já que as porções que mais contribuíram para a classificação serão realçadas.

5.3 Publicações Realizadas

O seguinte trabalho foi resultante das metodologias propostas e foi aceito para apresentação em uma conferência nacional:

1. JÚNIOR, E. G.; PAES, G.; SILVA, P. Classificação de tumor cerebral utilizando *deep learning*. Em Anais de IX Escola Regional de Computação Aplicada à Saúde. Porto Alegre, RS, Brasil: SBC, 2024. p. 29-32.

Referências

ABHRANTA PANIGRAH. Brain Tumor Detection MRI. 2021. Disponível em: <https://www.kaggle.com/datasets/abhranta/brain-tumor-detection-mri>. Acessado em 30 de Abril de 2024.

ALRABAI, A. Detection and diagnosis of brain tumor using convolutional neural networks. International Journal of Engineering Research, v. 2, n. 2, p. 26–34, 2023.

AYZENBERG, L.; GIRYES, R.; GREENSPAN, H. Dinov2 based self supervised learning for few shot medical image segmentation. arXiv preprint arXiv:2403.03273, 2024.

AZEVEDO, L. d. S. C. d. Detecção de tumor cerebral a partir de análise de imagens médicas usando inteligência artificial. Universidade Federal de Uberlândia, 2023.

BLOG, K. Métricas de Avaliação em Machine Learning: Classificação. 2020. Disponível em: <https://medium.com/kunumi/m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-em-machine-learning-classifica%C3%A7%C3%A3o-49340dcdb198>. Acessado em 21 de Maio de 2024.

BONDY, M. L.; SCHEURER, M. E.; MALMER, B.; BARNHOLTZ-SLOAN, J. S.; DAVIS, F. G.; IL'YASOVA, D.; KRUCHKO, C.; MCCARTHY, B. J.; RAJARAMAN, P.; SCHWARTZBAUM, J. A. et al. Brain tumor epidemiology: consensus from the brain tumor epidemiology consortium. Cancer, Wiley Online Library, v. 113, n. S7, p. 1953–1968, 2008.

BUCHANAN, B. G. A (Very) Brief History of Artificial Intelligence. 2006. Disponível em: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1848>. Acessado em 21 de Abril de 2024.

CASELI, H.; FREITAS, C.; VIOLA, R. Processamento de linguagem natural. Sociedade Brasileira de Computação, 2022.

CASPER, S.; LI, Y.; LI, J.; BU, T.; ZHANG, K.; HARIHARAN, K.; HADFIELD-MENELL, D. Red Teaming Deep Neural Networks with Feature Synthesis Tools. 2023. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2023/file/febe5c5c6973f713cc43bf0f7c90edbe-Paper-Conference.pdf>. Acessado em 21 de Abril de 2024.

CHAO, W.-L.; HU, H.; SHA, F. Cross-dataset adaptation for visual question answering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2018. p. 5716–5725.

CHARLES, N. A.; HOLLAND, E. C.; GILBERTSON, R.; GLASS, R.; KETTENMANN, H. The brain tumor microenvironment. Glia, Wiley Online Library, v. 59, n. 8, p. 1169–1180, 2011.

CHEN, S.; HOU, W.; KHAN, S.; KHAN, F. S. Progressive semantic-guided vision transformer for zero-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [s.n.], 2024. p. 31784. Accessed: December 2024. Disponível em: <https://arxiv.org/abs/2404.07713>.

CLARE, C. Cidade dos Ossos. [S.l.]: Galera Record, 2010.

COELHO, M. (Imagem) Exemplo de rede neural artificial. 2017. Disponível em: <http://www2.decom.ufop.br/imobilis/wp-content/uploads/2017/06/neuralNetwork.png>. Acessado em 09 de Abril de 2024.

CUNHA, J. P. Z. Um estudo comparativo das técnicas de avaliação cruzada aplicadas a modelos mistos. 2019. Disponível em: <https://www.teses.usp.br/teses/disponiveis/45/45133/tde-26082019-220647/publico/Dissertacao_JoaoPauloZanola.pdf>. Acessado em 30 de Abril de 2024.

DAMM, S.; LASZKIEWICZ, M.; LEDERER, J.; FISCHER, A. Anomalydino: Boosting patch-based few-shot anomaly detection with dinov2. arXiv preprint arXiv:2405.14529, 2024.

DAVIES, J.; SHERRIFF, N. Assessing public health policy approaches to level-up the gradient in health inequalities: the gradient evaluation framework. Public Health, Elsevier, v. 128, n. 3, p. 246–253, 2014.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2009. p. 248–255.

DOSOVITSKIY. Vision Transformer. 2024. Disponível em: <https://paperswithcode.com/method/vision-transformer>. Acessado em 05 de Agosto de 2024.

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEHGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S.; USZKOREIT, J.; HOULSBY, N. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020. Disponível em: <https://arxiv.org/abs/2010.11929>.

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEHGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S. et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEHGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S.; USZKOREIT, J.; HOULSBY, N. An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR). [s.n.], 2021. Disponível em: <https://openreview.net/forum?id=YicbFdNTTy>.

FERNANDES, F. M. L.; COSTA, L. P.; ANDRADE, S. M. M. dos S.; NETO, J. M. R. de S.; SILVA, J. H. B. da; VILLANUEVA, J. M. M. Estudo avaliativo dos biomarcadores mais influentes para doença de alzheimer com k-means e floresta aleatória. In: Congresso Brasileiro de Automática-CBA. [S.l.: s.n.], 2022. v. 3, n. 1.

GAO, Y.; SHI, C.; SONG, R. Deep spectral q-learning with application to mobile health. Stat, Wiley Online Library, v. 12, n. 1, p. e564, 2023.

GARCIA, S. C. O uso de árvores de decisão na descoberta de conhecimento na área da saúde. 2003.

HALDAR, N. A. H.; KHAN, F. A.; ALI, A.; ABBAS, H. Arrhythmia classification using mahalanobis distance based improved fuzzy c-means clustering for mobile health monitoring systems. Neurocomputing, Elsevier, v. 220, p. 221–235, 2017.

HAVAEI, M.; DAVY, A.; WARDE-FARLEY, D.; BIARD, A.; COURVILLE, A.; BENGIO, Y.; PAL, C.; JODOIN, P.-M.; LAROCHELLE, H. Deep learning for brain tumor classification and segmentation using mri images. Brain Tumor Segmentation with Deep Learning, Springer, v. 1, n. 1, p. 1–12, 2017.

HE, X.; PENG, Y. Fine-grained visual-textual representation learning. IEEE Transactions on Circuits and Systems for Video Technology, Institute of Electrical and Electronics Engineers (IEEE), v. 30, n. 2, p. 520–531, fev. 2020. ISSN 1558-2205. Disponível em: <http://dx.doi.org/10.1109/TCSVT.2019.2892802>.

HUANG, Y.; ZOU, J.; MENG, L.; YUE, X.; ZHAO, Q.; LI, J.; SONG, C.; JIMENEZ, G.; LI, S.; FU, G. Comparative analysis of imagenet pre-trained deep learning models and dinov2 in medical imaging classification. arXiv preprint arXiv:2402.07595, 2024.

IBM. O que é machine learning? 2023. Disponível em: <https://www.ibm.com/br-pt/topics/machine-learning>. Acessado em 20 de Abril de 2024.

JúNIOR, E. G.; PAES, G.; SILVA, P. Classificação de tumor cerebral utilizando deep learning. In: Anais da IX Escola Regional de Computação Aplicada à Saúde. Porto Alegre, RS, Brasil: SBC, 2024. p. 29–32. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/ercas/article/view/29692>.

KHAN, M. K. H.; GUO, W.; LIU, J.; DONG, F.; LI, Z.; PATTERSON, T. A.; HONG, H. Machine learning and deep learning for brain tumor mri image segmentation. Experimental Biology and Medicine, SAGE Publications Sage UK: London, England, v. 248, n. 21, p. 1974–1992, 2023.

KIM, J.; SHIM, K.; KIM, J.; SHIM, B. Vision Transformer-based Feature Extraction for Generalized Zero-Shot Learning. 2023. Disponível em: <https://arxiv.org/abs/2302.00875>.

KUFEL, J.; BARGIEŁ-ŁĄCZEK, K.; KOCOT, S.; KOŹLIK, M.; BARTNIKOWSKA, W.; JANIK, M.; CZOGALIK, Ł.; DUDEK, P.; MAGIERA, M.; LIS, A. et al. What is machine learning, artificial neural networks and deep learning?—examples of practical applications in medicine. Diagnostics, MDPI, v. 13, n. 15, p. 2582, 2023.

LAMRANI, D.; CHERRADI, B.; GANNOUR, O. E.; BOUQENTAR, M. A.; BAHATTI, L. Brain tumor detection using mri images and convolutional neural network. International Journal of Advanced Computer Science and Applications, Science and Information (SAI) Organization Limited, v. 13, n. 7, 2022.

LITJENS, G.; CIOMPI, F.; GHAFOORIAN, M.; BULTEN, W.; MIESENBERGER, K.; GONçALVES, A.; ZUIDHOF, G.; SáNCHEZ, C. I. Deep learning in radiology: An overview of the concepts and a survey of the state of the art. Journal of Magnetic Resonance Imaging, Wiley Online Library, v. 46, n. 3, p. 872–879, 2017.

LIU, Y.; PU, H.; SUN, D.-W. Efficient extraction of deep image features using convolutional neural network (cnn) for applications in detecting and analysing complex food matrices. Trends in Food Science & Technology, Elsevier, v. 113, p. 193–204, 2021.

MAGNO, L. Matriz de Confusão: nunca mais se confunda utilizando esse exemplo. 2023. Disponível em: <https://medium.com/comunidadeds/matriz-de-confus%C3%A3o-nunca-mais-se-confunda-utilizando-esse-exemplo-35a9ac63b88a>. Acessado em 21 de Maio de 2024.

Marina Baeta. Tumores cerebrais: uma revisão. 2020. Disponível em: <https://sanarmed.com/tumores-cerebrais-uma-revisao-colunistas/>. Acessado em 17 de Abril de 2024.

MAURÍCIO, J.; DOMINGUES, I.; BERNARDINO, J. Comparing vision transformers and convolutional neural networks for image classification: A literature review. Applied Sciences, MDPI, v. 13, n. 9, p. 5521, 2023.

MAYANGSARI, M. K.; SYARIF, I.; BARAKBAH, A. Evaluation of stratified k-fold cross validation for predicting bug severity in game review classification. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 2023.

MORAES, J. R. d.; MOREIRA, J. P. d. L.; LUIZ, R. R. Associação entre o estado de saúde autorreferido de adultos e a área de localização do domicílio: uma análise de regressão logística ordinal usando a pnad 2008. Ciência & saúde coletiva, SciELO Public Health, v. 16, p. 3769–3780, 2011.

MOUTIK, O.; SEKKAT, H.; TIGANI, S.; CHEHRI, A.; SAADANE, R.; TCHAKOUCHT, T. A.; PAUL, A. Convolutional neural networks or vision transformers: Who will win the race for action recognitions in visual data? Sensors, MDPI, v. 23, n. 2, p. 734, 2023.

NALEPA, J.; MARCINKIEWICZ, M.; KAWULOK, M. Data augmentation for brain-tumor segmentation: a review. Frontiers in computational neuroscience, Frontiers Media SA, v. 13, p. 83, 2019.

Navoneel Chakrabarty. Brain MRI Images for Brain Tumor Detection. 2019. Disponível em: <https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection>. Acessado em 06 de Janeiro de 2024.

NETO, G. R. L. Aplicação de machine learning para classificação de imagens astronômicas. Universidade Estadual Paulista (Unesp), 2023.

OQUAB, M.; DARCET, T.; MOUTAKANNI, T.; VO, H.; SZAFRANIEC, M.; KHALIDOV, V.; FERNANDEZ, P.; HAZIZA, D.; MASSA, F.; EL-NOUBY, A. et al. Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193, 2023.

OQUAB, M.; DARCET, T.; MOUTAKANNI, T.; VO, H.; SZAFRANIEC, M.; KHALIDOV, V.; FERNANDEZ, P.; HAZIZA, D.; MASSA, F.; EL-NOUBY, A. et al. Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193, 2023.

OQUAB, T. D. M. DINOv2: Learning Robust Visual Features without Supervision. 2024. Disponível em: <https://ar5iv.labs.arxiv.org/html/2304.07193>. Acessado em 13 de Junho de 2024.

PARK, J.; PARK, Y. G. Brain tumor rehabilitation: symptoms, complications, and treatment strategy. Brain & Neurorehabilitation, Korean Society for NeuroRehabilitation, v. 15, n. 3, 2022.

PEREIRA, M. F. S.; LEITE, C. A.; OLIVEIRA, L. S.; LIMA, A. A. de M.; ANELLI, R.; MARTINS, A. F. M. A survey on deep learning in medical image analysis. Journal of Health Informatics, Universidade Estadual Paulista (UNESP), v. 11, n. 3, p. 120–136, 2019.

POURPANAH, F.; ABDAR, M.; LUO, Y.; ZHOU, X.; WANG, R.; LIM, C. P.; WANG, X.-Z.; WU, Q. J. A review of generalized zero-shot learning methods. IEEE transactions on pattern analysis and machine intelligence, IEEE, v. 45, n. 4, p. 4051–4070, 2022.

RAGHU, M.; UNTERTHINER, T.; KORNBLITH, S.; ZHANG, C.; DOSOVITSKIY, A. Do Vision Transformers See Like Convolutional Neural Networks? 2022. Disponível em: <https://arxiv.org/abs/2108.08810>.

RANFTL, R.; LASINGER, K.; HAFNER, D.; SCHINDLER, K.; KOLTUN, V. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. IEEE transactions on pattern analysis and machine intelligence, IEEE, v. 44, n. 3, p. 1623–1637, 2020.

RUNDO, L.; HAN, C.; ZHANG, J.; HATAYA, R.; NAGANO, Y.; MILITELLO, C.; FERRETTI, C.; NOBILE, M. S.; TANGHERLONI, A.; GILARDI, M. C. et al. Cnn-based prostate zonal segmentation on t2-weighted mr images: a cross-dataset study. Neural Approaches to Dynamics of Signal Exchanges, Springer, p. 269–280, 2020.

RUNSTEIN, F. O. Sistema de reconhecimento de fala baseado em redes neurais artificiais. Tese (Doutorado) — University of Campinas, Brazil, 1998.

SAKURAI, R. Implementando a estrutura de uma Rede Neural Convolucional utilizando o MapReduce do Spark. 2017. Disponível em: <https://www.sakurai.dev.br/cnn-mapreduce/>. Acessado em 17 de Maio de 2024.

SAVEGNAGO, G. D. O.; PINTO, G. V.; SNOVARESK, C. F.; HAMAD, N. de O.; SERPA, G. F.; LIEDKE, G. S. Inteligência artificial na odontologia: uma revisão narrativa de literatura. 2024. Disponível em: <https://seer.upf.br/index.php/rfo/article/view/15733/114117855>. Acessado em 21 de Abril de 2024.

SHABBIR, A.; NAZIR, K. et al. Brain tumor detection based on deep learning approach. Journal of Computing & Biomedical Informatics, v. 4, n. 02, p. 298–310, 2023.

SHEWAN, D. 10 Companies Using Machine Learning in Cool Ways. 2023. Disponível em: <https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications>. Acessado em 29 de Abril de 2024.

SILVA, P.; LUZ, E.; SILVA, G.; MOREIRA, G.; SILVA, R.; LUCIO, D.; MENOTTI, D. Covid-19 detection in ct images with deep learning: A voting-based scheme and cross-datasets analysis. Informatics in medicine unlocked, Elsevier, v. 20, p. 100427, 2020.

SILVA, R. E. V. d. Um estudo comparativo entre redes neurais convolucionais para a classificação de imagens. 2018.

SOARES, R. A.; PEREIRA, I. S.; FRAZÃO, M. P.; DUQUE, M. d. G. C.; DUQUE, R. d. G. C.; PÁDUA, D. M.; MARTINS, J. K. G. da R.; PEIXOTO, J. de O.; ACÁCIO, M. da S.; GALVÃO, A. A. C. B. et al. O uso da inteligência artificial na medicina: aplicações e benefícios. Research, Society and Development, v. 12, n. 4, p. e5012440856–e5012440856, 2023.

SOORI, M.; AREZOO, B.; DASTRES, R. Machine learning and artificial intelligence in cnc machine tools, a review. Sustainable Manufacturing and Service Economics, Elsevier, v. 2, p. 100009, 2023.

STRONG, A. Applications of artificial intelligence & associated technologies. Science [ETEBMS-2016], v. 5, n. 6, p. 64–67, 2016.

SUN, C.; SHRIVASTAVA, A.; SINGH, S.; GUPTA, A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. 2017. Disponível em: <https://arxiv.org/abs/1707.02968>.

SUNIGA, A. Conjuntos de treino, validação e teste em Machine Learning. 2020. Disponível em: <https://medium.com/@abnersuniga7/conjuntos-de-treino-teste-e-valida%C3%A7%C3%A3o-em-machine-learning-fast-ai-5da612dcb0ed>. Acessado em 01 de Agosto de 2024.

TAJBAKHSH, N.; SHIN, J. H.; GURUDU, S.; HURST, R. T.; KENDALL, C. B. Application of deep learning in neuroimaging: A systematic review. Journal of Neuroimaging, Wiley Online Library, v. 30, n. 4, p. 387–398, 2020.

TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. International conference on machine learning. [S.l.], 2019. p. 6105–6114.

TAUD, H.; MAS, J.-F. Multilayer perceptron (mlp). Geomatic approaches for modeling land change scenarios, Springer, p. 451–455, 2018.

ULLAH, N.; JAVED, A.; ALHAZMI, A.; HASNAIN, S. M.; TAHIR, A.; ASHRAF, R. Tumordetnet: A unified deep learning model for brain tumor detection and classification. Plos one, Public Library of Science San Francisco, CA USA, v. 18, n. 9, p. e0291200, 2023.

WU, B.; XU, C.; DAI, X.; WAN, A.; ZHANG, P.; YAN, Z.; TOMIZUKA, M.; GONZALEZ, J.; KEUTZER, K.; VAJDA, P. Visual Transformers: Token-based Image Representation and Processing for Computer Vision. 2020.

ZHOU, J.; WEI, C.; WANG, H.; SHEN, W.; XIE, C.; YUILLE, A.; KONG, T. iBOT: Image BERT Pre-Training with Online Tokenizer. 2022. Disponível em: <https://arxiv.org/abs/2111.07832>.