UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE ESPECIALIZAÇÃO EM ENGENHARIA DE SOFTWARE E CIÊNCIA DE DADOS

LUCAS JOSÉ PEREIRA DA COSTA

Avaliação do Desempenho de Arquiteturas de Deep Learning na Classificação de Imagens Médicas

Monografia apresentada como requisito parcial para a obtenção do grau de Especialista em Engenharia de Software e Ciência de Dados.

Orientador: Prof. Dr. Anderson Rocha Tavares

Porto Alegre
2026

CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Costa, Lucas José Pereira da

Avaliação do Desempenho de Arquiteturas de Deep Learning na Classificação de Imagens Médicas / Lucas José Pereira da Costa. – 2026.

57 f.

Orientador: Anderson Rocha Tavares.

Trabalho de conclusão de curso (Especialização) – Universidade Federal do Rio Grande do Sul, Instituto de Informática, Curso de Especialização em Engenharia de Software e Ciência de Dados, Porto Alegre, BR-RS, 2026.

1. Doença de Alzheimer. 2. Imagens por Ressonância Magnética. 3. Aprendizado Profundo (Deep Learning). 4. Redes Neurais Convolucionais. I. Tavares, Anderson Rocha, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

“If I have seen farther than others,
it is because I stood on the shoulders of giants.”
— Sir Isaac Newton

Agradecimentos

À minha esposa, Lívia, por toda a paciência, apoio incondicional e compreensão demonstrados durante as longas horas dedicadas à realização deste trabalho. Sua presença foi fundamental para que eu pudesse superar os desafios desta jornada.

Ao meu orientador, Prof. Dr. Anderson Rocha Tavares, pela orientação precisa, disponibilidade e pelos valiosos ensinamentos que moldaram o rigor científico deste projeto.

Aos professores da Universidade Federal do Rio Grande do Sul (UFRGS), que contribuíram para a minha formação acadêmica e profissional ao longo de todo o curso.

Nota de Agradecimento aos Dados: Os dados utilizados neste trabalho foram fornecidos pelo projeto OASIS-1: *Principal Investigators: D. Marcus, R. Buckner, J. Csernansky, J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382*.

Resumo

Este trabalho avalia o desempenho de diferentes arquiteturas de Deep Learning aplicadas à classificação de imagens médicas tridimensionais para detecção da Doença de Alzheimer, utilizando o conjunto de dados OASIS. Foram comparadas abordagens 2D e 3D. Modelos 2D como ResNet18 e AlexNet apresentaram resultados consistentes mesmo processando apenas slices individuais, mas essa estratégia perde informações volumétricas importantes. Também foram testadas CNNs 3D, que capturam melhor a estrutura do cérebro, mas tendem ao overfitting devido ao tamanho reduzido do dataset. Técnicas de interpretabilidade, como Grad-CAM, foram aplicadas para verificar se regiões relevantes, como o hipocampo, contribuíam para a decisão das redes. Os resultados mostram que arquiteturas leves alcançam desempenho competitivo em ambientes com poucos dados, reforçando a importância de soluções otimizadas e interpretáveis para análise de imagens médicas.

Abstract

This study evaluates the performance of different Deep Learning architectures applied to the classification of three-dimensional medical images for the detection of Alzheimer's Disease, using the OASIS dataset. Both 2D and 3D approaches were compared. 2D models such as ResNet18 and AlexNet showed consistent results even when processing individual slices, although this strategy loses important volumetric information. 3D CNNs were also tested, which better capture the brain's structure but tend to overfit due to the small dataset size. Interpretability techniques, such as Grad-CAM, were applied to verify whether relevant regions, such as the hippocampus, contributed to the networks' decisions. The results indicate that lightweight architectures achieve competitive performance in data-limited environments, highlighting the importance of optimized and interpretable solutions for medical image analysis.

LISTA DE ABREVIATURAS E SIGLAS

ASIC Circuito Integrado de Aplicação Específica (do inglês *Application-Specific Integrated Circuit*)

CAM *Class Activation Map*

CNN *Convolutional Neural Network*

DA Doença de Alzheimer

FC Totalmente Conectada (do inglês, *Fully Connected*)

FPGA *Field-Programmable Gate Array*

GAP Agrupamento Médio Global (do inglês, *Global Average Pooling*)

IA Inteligência Artificial

LSTM *Long Short-Term Memory*

MNIST *Modified National Institute of Standards and Technology*

ReLU *Rectified Linear Unit*

ResNet Redes Residuais Profundas (do inlês, *Residual Neural Network*)

SIANN *Shift-Invariant Artificial Neural Network*

Lista de Figuras

Figura 2.1 Imagem de um cérebro saudável versus um com demência, adaptado de (FILHO; VIDAL, 2024). 17
Figura 2.2 Os estágios do Alzheimer, da esquerda para a direita, são: normal, leve e grave (AWANG; ALI; FAHEEM, 2024). 17
Figura 2.3 Fatias coronais de uma ressonância magnética de um paciente com doença de Alzheimer. (BRUNO; SIQUEIRA et al., 2022). 17
Figura 2.4 Representação conceitual de uma rede neural. Fonte (FILHO; VIDAL, 2024). 19
Figura 2.5 Bloco residual de uma ResNet. Fonte: (FEITOSA, 2024) 21
Figura 2.6 Fonte (FEITOSA, 2024) 26

Figura 4.1 Distribuição da quantidade de imagens por classe diagnóstica no dataset OASIS-1. 35
Figura 4.2 Eixos de análise retirado de (FILHO; VIDAL, 2024). 35
Figura 4.3 Distribuição dos dados por paciente e separação em treino, teste e validação. 36

Figura 5.1 Matriz de Confusão *AlexNet* 42
Figura 5.2 Matriz de Confusão *SqueezeNet* 42
Figura 5.3 Matriz de confusão *ResNet18* 43
Figura 5.4 Matriz de confusão *EfficienteNet-B3* 43
Figura 5.5 Mapa Grad-CAM do modelo *AlexNet* 44
Figura 5.6 Mapa Grad-CAM do modelo *SqueezeNet* 45
Figura 5.7 Mapa Grad-CAM do modelo *ResNet18* 45
Figura 5.8 Mapa Grad-CAM do modelo *EfficientNet-B3* 46
Figura 5.9 Matix de confusão do modelo *Conv3D* 47
Figura 5.10 Classe verdadeira: *Non Demented*, Predição: *Non Demented* 48
Figura 5.11 Classe verdadeira: *Very mild Dementia*, Predição: *Very mild Dementia* 49
Figura 5.12 Classe verdadeira: *Non Demented*, Predição: *Very mild Dementia* 50
Figura 5.13 Matriz de confusão do Modelo *Conv3d* 51

Lista de Tabelas

Tabela 4.1 Hiperparâmetros utilizados nos experimentos. 38

Tabela 5.1 Comparativo de métricas de desempenho: *AlexNet*, *SqueezeNet* , *ResNet-18* e *EfficientNet-B3*. 39
Tabela 5.2 Comparativo de métricas de desempenho: *EfficientNet-B3* em diferentes resoluções. 40
Tabela 5.3 Histórico de treino dos modelos *AlexNet*, *SqueezeNet*, *ResNet-18* e *EfficientNet-B3*. 41

Sumário

1 INTRODUÇÃO 12
1.1 Justificativa 14
1.2 Objetivos 14
1.2.1 Objetivo Geral 14
1.2.2 Objetivos Específicos 15
2 FUNDAMENTOS 16
2.1 Visão Geral da Doença de Alzheimer 16
2.2 Neuroimagem 18
2.3 Redes Neurais 18
2.4 Redes Neurais Convolucionais 20
2.4.1 Redes Residuais Profundas 20
2.4.2 SqueezeNet 22
2.4.3 EfficientNet 23
2.4.4 Conv3D 23
2.5 Interpretabilidade em Inteligência Artificial 24
2.5.1 Grad-CAM 25
2.5.1.1 Layer Grad-CAM 26
2.5.2 Mapa de Saliência 26
3 TRABALHOS RELACIONADOS 28
3.1 Metodologia do Levantamento Bibliográfico 28
3.2 Resultados 30
3.2.1 Q1: Métodos de explicabilidade em modelos de visão computacional aplicados a imagens médicas 30
3.2.2 Q2: Arquiteturas de redes neurais mais utilizadas com técnicas de explicabilidade na detecção de Alzheimer 30
3.2.3 Q3: Limitações das técnicas de explicabilidade em imagens médicas 31
3.3 Resumo 31
4 METODOLOGIA 32
4.1 Modelos de CNN Pré-Treinados 32
4.2 Arquitetura Convolucional Tridimensional (Conv3D) 33
4.2.1 Redes Neurais Utilizadas 34
4.2.2 Dataset Utilizado 34
4.2.3 Divisão do Conjunto de Dados por Paciente 36
4.3 Configurações Experimentais 37
4.4 Grad-CAM 38
5 RESULTADOS 39
5.1 Resultados da Execução dos Modelos CNN 2D 39
5.2 Conv3D 47
5.3 Resumo 51
6 CONCLUSÃO 52

REFERÊNCIAS.....54

1 Introdução

O diagnóstico por imagem é uma das ferramentas mais relevantes da medicina contemporânea, sendo indispensável na detecção e no acompanhamento de diversas enfermidades. Exames como raio-X, tomografia computadorizada e ressonância magnética permitem a visualização precisa das estruturas internas do corpo humano, desempenhando papel crucial na rotina clínica. Contudo, a análise dessas imagens depende da expertise dos profissionais de saúde, o que pode gerar variações interpretativas e, eventualmente, ocasionar atrasos ou equívocos (WAITE et al., 2017; GEGENFURTNER et al., 2017). Nesse cenário, técnicas de inteligência artificial, especialmente aquelas baseadas em *Deep Learning*, têm se mostrado promissoras ao oferecer suporte à tomada de decisões médicas, elevando os níveis de precisão e agilidade na interpretação dos exames (LI et al., 2023; AGGARWAL et al., 2021). Os estágios iniciais da doença de Alzheimer são frequentemente confundidos com alterações cognitivas naturais do envelhecimento, dificultando o diagnóstico precoce. Além disso, a análise de imagens médicas pode envolver subjetividade e depender fortemente da experiência do especialista. Nesse contexto, soluções computacionais baseadas em ciência de dados e aprendizado de máquina tornam-se importantes para apoiar o diagnóstico precoce. Embora diversas abordagens utilizem imagens de ressonância magnética cerebral com resultados satisfatórios, muitas dependem de etapas de pré-processamento específicas e fortemente baseadas na anatomia cerebral, o que pode limitar sua generalização e reprodutibilidade (PADOVESE, 2017). A crescente demanda por soluções computacionais eficientes, capazes de aprimorar a qualidade dos serviços de saúde, motiva esta pesquisa. Dentre as abordagens mais eficazes, destacam-se as redes neurais profundas — em particular, as convolucionais (CNNs) e os modelos recentes baseados em *transformers* — que se sobressaem na identificação de padrões complexos em grandes conjuntos de dados. Essas arquiteturas vêm obtendo resultados expressivos em tarefas como classificação, segmentação e detecção de anomalias em imagens médicas (TAKAHASHI et al., 2024; MÜLLER; SOTO-REY; KRAMER,

2022). Ainda assim, persistem desafios significativos, como a escassez de dados rotulados, o desbalanceamento entre categorias e a elevada demanda computacional para treinar modelos robustos (BARGAGNA et al., 2023).

Nesse contexto, abordagens tridimensionais têm recebido destaque. O *deep learning*, especialmente por meio das redes neurais convolucionais tridimensionais (3D CNNs), tem se mostrado eficaz em diversas tarefas de classificação de imagens médicas devido à sua capacidade de extrair automaticamente características a partir de dados volumétricos. Entretanto, o treinamento de modelos 3D CNN exige grandes quantidades de dados, o que se torna um desafio em domínios como a análise de ressonância magnética para detecção da doença de Alzheimer, onde os conjuntos disponíveis frequentemente são limitados, aumentando a probabilidade de *overfitting*. Por esse motivo, técnicas de *data augmentation* têm sido amplamente empregadas para gerar artificialmente amostras adicionais, contribuindo para melhorar a acurácia e a capacidade de generalização dos modelos (VO; HO, 2025).

Além disso, embora a ressonância magnética estrutural revele biomarcadores precoces, como a atrofia do hipocampo, métodos convencionais baseados em CNNs 2D tendem a perder informações volumétricas relevantes, enquanto abordagens 3D de cérebro inteiro enfrentam desafios relacionados à alta complexidade computacional e ao risco de *overfitting* (TURANLI, 2025). Esses aspectos reforçam a importância de comparar arquiteturas distintas para identificar modelos que conciliem eficiência computacional e desempenho.

Diante dessas questões, o presente trabalho propõe uma análise comparativa entre diferentes arquiteturas de redes neurais profundas aplicadas à classificação de imagens médicas. A pesquisa utilizará bases públicas de exames radiológicos para testar modelos consolidados — como *AlexNet*, *ResNet18* e *SqueezeNet* — além de modelos volumétricos como *3D CNNs*. A avaliação abrangerá não apenas métricas de desempenho, como acurácia, mas também fatores como tempo de treinamento e uso de recursos computacionais, essenciais em cenários com infraestrutura limitada.

O estudo também visa compreender os efeitos do desbalanceamento de classes na performance dos modelos, propondo estratégias de correção como reamostragem e a adoção de métricas mais adequadas. Os resultados esperados pretendem subsidiar o desenvolvimento de ferramentas automatizadas de apoio ao diagnóstico, contribuindo para ampliar o acesso à saúde de qualidade, especialmente em regiões com escassez de profissionais especializados.

Assim, esta proposta busca aliar avanços tecnológicos à aplicação prática em um dos setores mais sensíveis da sociedade, oferecendo soluções que dialogam com as necessidades da saúde pública.

1.1 Justificativa

A escolha de avaliar arquiteturas de Deep Learning para a classificação de imagens médicas está alinhada à crescente demanda por soluções computacionais que conciliem precisão diagnóstica e eficiência no processamento de dados. A relevância deste estudo se reforça diante de cenários em que o acesso a hardware de alto desempenho é limitado, como ocorre em muitas instituições de saúde. Nesta pesquisa, serão utilizados recursos de GPU disponibilizados pelo Google Colab e pelo Kaggle, devido à ausência de hardware dedicado. Essa limitação impõe restrições à execução dos experimentos, como maior tempo de treinamento e a necessidade de ajustes nos tamanhos de lote e na complexidade dos modelos.

1.2 Objetivos

Este trabalho tem como principal finalidade investigar o desempenho de diferentes arquiteturas de *Deep Learning* na tarefa de classificação de imagens médicas, especialmente aquelas provenientes de exames como a ressonância magnética.

1.2.1 Objetivo Geral

Avaliar e comparar o desempenho de modelos baseados em redes neurais convolucionais (CNNs) e transformers na classificação de imagens médicas, com foco em métricas como acurácia, precisão, revocação e *F1-score*, utilizando bases de dados públicas.

1.2.2 Objetivos Específicos

- Realizar o pré-processamento e a análise exploratória das bases de dados públicas de imagens médicas utilizadas no estudo;
- Investigar e aplicar estratégias para mitigação do desbalanceamento entre classes, visando melhorar a capacidade de generalização dos modelos;
- Desenvolver e implementar modelos de classificação de imagens médicas baseados em arquiteturas de Redes Neurais Convolucionais (CNNs), incluindo abordagens 2D (AlexNet, ResNet18, SqueezeNet e EfficientNet-B3) e abordagem 3D (Conv3D);
- Avaliar o desempenho dos modelos por meio de métricas quantitativas adequadas ao contexto médico, analisando sua eficiência e capacidade de generalização;
- Comparar os resultados obtidos entre as diferentes arquiteturas, identificando vantagens, limitações e oportunidades de melhoria;
- Investigar técnicas de processamento de imagens médicas que contribuam para a melhoria da qualidade dos dados e do desempenho dos modelos.

Este trabalho está organizado em seis capítulos, estruturados de modo a apresentar, de forma progressiva, os fundamentos teóricos, a metodologia adotada e os resultados obtidos.

No Capítulo 2, são apresentados os fundamentos teóricos sobre a Doença de Alzheimer, redes neurais convolucionais e técnicas de interpretabilidade aplicadas à análise de imagens médicas.

O Capítulo 3 aborda os principais métodos e arquiteturas encontrados na literatura, com foco nas técnicas de explicabilidade em modelos de visão computacional voltados à área médica.

No Capítulo 4, descreve-se a metodologia adotada, incluindo o pré-processamento dos dados, a divisão do conjunto por paciente, as arquiteturas utilizadas, bem como as estratégias de treinamento e avaliação.

No Capítulo 5, são apresentados e discutidos os resultados experimentais, por meio das métricas de desempenho, matrizes de confusão e mapas Grad-CAM, destacando as limitações observadas.

Por fim, o Capítulo 6 apresenta as conclusões do trabalho e as perspectivas para pesquisas futuras.

2 Fundamentos

Este capítulo apresenta os fundamentos teóricos que sustentam o desenvolvimento deste trabalho. Inicialmente, é realizada uma contextualização da Doença de Alzheimer, destacando suas principais características clínicas e estruturais, bem como sua relevância no cenário da saúde pública. Em seguida, são introduzidos os conceitos de neuroimagem e representação digital de imagens médicas, com ênfase na ressonância magnética. Posteriormente, são discutidos os princípios das redes neurais artificiais e das redes neurais convolucionais, incluindo arquiteturas relevantes e métodos de interpretabilidade, estabelecendo a base conceitual necessária para a aplicação de técnicas de aprendizado profundo na análise de imagens cerebrais.

2.1 Visão Geral da Doença de Alzheimer

A demência pode ser caracterizada como um conjunto de condições que comprometem a memória, a linguagem ou o raciocínio lógico do indivíduo. A causa mais comum desse quadro é a Doença de Alzheimer (DA), responsável por aproximadamente 60% a 80% dos casos de demência (FILHO; VIDAL, 2024). A doença evolui de forma gradual e não possui cura; geralmente, indivíduos diagnosticados com DA apresentam expectativa de vida entre 4 e 8 anos, podendo chegar a até 20 anos, dependendo de diversos fatores (FILHO; VIDAL, 2024).

A Figura 2.1 ilustra visualmente as implicações da Doença de Alzheimer, destacando as áreas afetadas e as alterações estruturais ao longo de sua progressão. Embora a atrofia seja um processo natural do envelhecimento, o nível de encolhimento observado na imagem à direita evidencia uma perda acentuada de neurônios. Essa perda pode estar associada ao acúmulo de placas amiloides e emaranhados neurofibrilares, que contribuem para a morte celular. O hipocampo (indicado por setas azuis) também apresenta sinais de comprometimento (FILHO; VIDAL, 2024).

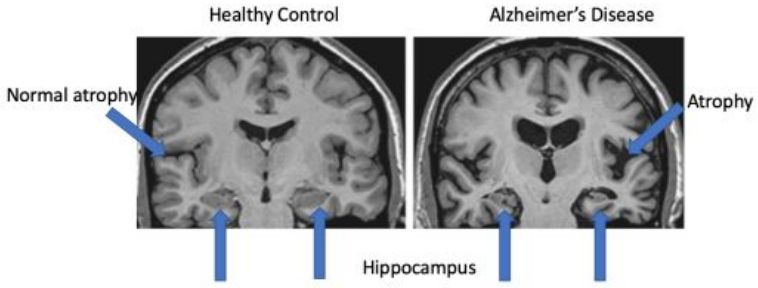


Figura 2.1 – Imagem de um cérebro saudável versus um com demência, adaptado de (FILHO; VIDAL, 2024).



Figura 2.2 – Os estágios do Alzheimer, da esquerda para a direita, são: normal, leve e grave (AWANG; ALI; FAHEEM, 2024).

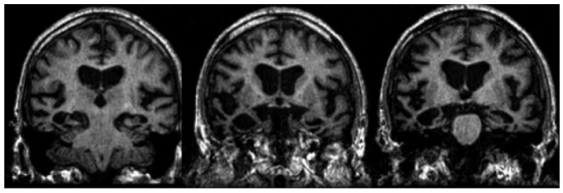


Figura 2.3 – Fatias coronais de uma ressonância magnética de um paciente com doença de Alzheimer. (BRUNO; SIQUEIRA et al., 2022).

A Figura 2.3 mostra a redução acentuada do volume de ambos os hipocampos, mais intensa no lado direito (BRUNO; SIQUEIRA et al., 2022).

Uma maneira de explicar ludica de explicar um um exame de ressonia magnetica imagina que você tem um pão de forma se você cortar fatias bem finas (*slice*) você consegue identicar que alguma partes do pão estão mofadas por exemplo.

No caso da ressonacia o médico precisa anilisar varias imagens que pode ser um processo bastante demorado alem disso os aparelhos modernos de ressonancia maganicas consegue fazer slices cada vez mais finos e os radiologista precisam interpretar milhares de imagens (MCDONALD et al., 2015).

Essa compreensão das alterações estruturais é essencial para aprimorar o diagnóstico precoce, monitorar a progressão da doença e subsidiar o desenvolvimento de estratégias terapêuticas apoiadas por técnicas de Ciência de Dados.

2.2 Neuroimagem

Uma imagem digital pode ser considera uma representação de um objeto em formato eletronico. Esses objetos podem ser representados por matrizes de M x N, sendo que cada elemento dessa matriz representa um *pixel*, descrito pela Equação 2.1:

$$f(x, y) = I \tag{2.1}$$

onde *x* e *y* representam as coordenadas da matriz (linha e coluna) e I a intensidade de cor (JUNIOR, 2023).

No entanto, ao analisar os tipos de imagens que possuem dimensões além do plano, por exemplo, quando analisamos imagens cerebrais estamos interessados em uma imagem tridimencional completa do cérebro. Iso sifnifca que as imagens cerebrais são geralmente tridimensionais (JUNIOR, 2023). Os pesquisadores da área de neuroimagem utilizam imagens cerebrais obtidas por diferentes métodos de aquisição. No contexto deste trabalho, a ressonância magnética (RM) foi escolhida como técnica de análise devido à ampla disponibilidade de exames e à qualidade das informações fornecidas.

2.3 Redes Neurais

Em um primeiro momento, é importante compreender por que utilizar redes neurais convolucionais para a classificação de imagens de ressonância magnética e, consequentemente, como isso contribui para a identificação dos estágios da DA. Contudo, antes de abordar o conceito de convoluções, é fundamental entender o funcionamento das redes neurais e por que elas são relevantes nesse contexto.

O estudo das redes neurais surgiu como resposta aos desafios encontrados na classificação não linear. Com o objetivo de fornecer uma alternativa flexível e eficaz, as redes neurais operam por meio de camadas compostas por somas ponderadas e funções de ativação, que introduzem diretamente a não linearidade necessária ao modelo. A Figura 2.4 apresenta uma visão conceitual desse processo.

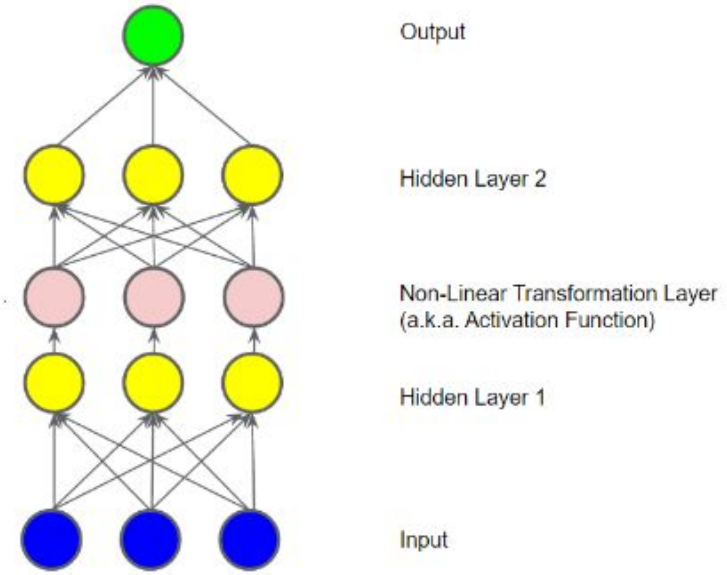


Figura 2.4 – Representação conceitual de uma rede neural. Fonte (FILHO; VIDAL, 2024).

À medida que mais camadas são adicionadas, a rede neural torna-se capaz de aprender e representar funções cada vez mais complexas a partir das entradas. Esse aumento de profundidade é um dos principais benefícios da arquitetura, pois amplia a capacidade da rede em compreender padrões e processar informações, permitindo representações significativamente mais sofisticadas dos dados de entrada (FILHO; VIDAL, 2024).

Existem duas funções de ativação muito conhecidas que são a *Sigmoid* (2.2) e a ReLU (2.3), observe a diferença delas de acordo com as suas equações:

$$F(x) = \frac{1}{1 + e^{-x}} \tag{2.2}$$

$$F(x) = \max(0, x) \tag{2.3}$$

Além de ser mais simples de calcular, a função ReLU (*Rectified Linear Unit*) costuma apresentar resultados superiores à função *Sigmoid*, pois seu gradiente é maior. Com isso, a partir desse entendimento inicial sobre redes neurais, pode-se avançar para o estudo das convoluções.

da uma revisada

2.4 Redes Neurais Convolucionais

Convolução no contexto de inteligencia artificial é a operação responsavel, com suporte das outras, em aprender as características das imagens. A camada de convolução aprende características de baixo nível como segmentos de retas, curvas, diagonais nas camadas iniciais, e posteriormente as utilizam para reconhecer caracteristicas de nível maior nas camadas seguintes (MARCHIORI et al., 2023).

As CNNs são arquiteturas de *Machine Learning* inspiradas no funcionamento do sistema visual biológico. Conforme descrito por (FEITOSA, 2024), células do córtex visual respondem a estímulos luminosos em campos receptivos, conceito que motivou o desenvolvimento do Neocognitron por Kunihiko Fukushima em 1980, uma rede hierárquica voltada ao reconhecimento de padrões.

Em 1989, a *LeNet-5* incorporou backpropagation em uma estrutura moderna para a classificação de dígitos manuscritos no dataset MNIST (LECUN et al., 2002).Na arquitetura LeNet-5, consiste no deslizamento de um filtro (ou kernel) sobre uma imagem de entrada para a extração de caracteristicas locais e a geração de mapas de ativação. A grande inovação da *LeNet-5* em relação ao processamento de imagens tradicional reside no fato de que os pesos desses filtros não são definidos manualmente para tarefas específicas, mas são parâmetros aprendíveis ajustados automaticamente via backpropagation durante o treinamento (LECUN et al., 2002). Na mesma linha, (ZHANG et al., 1996) propuseram a SIANN, uma rede neural invariante ao deslocamento para reconhecimento de caracteres. Contudo, essas abordagens iniciais ainda eram limitadas para tarefas mais complexas, como a classificação de imagens e vídeos de alta resolução.

2.4.1 Redes Residuais Profundas

À medida que as redes neurais se tornam mais profundas, compreender como a adição de camadas aumenta a complexidade e a capacidade expressiva do modelo torna-se crucial (ZHANG et al., 1996). Nesse contexto, em 2015, foi proposta a Rede Residual (*ResNet*) por (HE et al., 2016) para tarefas de reconhecimento de imagens. Essa arquitetura introduz conexões de atalho (shortcut connections), nas quais a entrada de um bloco funcional pula camadas intermediárias (geralmente duas ou três) para ser adicionada diretamente à saída posterior desse bloco. Embora amplamente difundida em estruturas convolucionais, a natureza residual é um princípio de design de rede que favorece o fluxo

de gradientes, mitigando o problema do desvanecimento do gradiente e contribuindo para a melhoria do desempenho em redes extremamente profundas (FEITOSA, 2024). A Figura 2.5 ilustra esse bloco residual, em que a linha sólida que transporta a entrada da camada cc para o operador de adição corresponde à conexão residual.

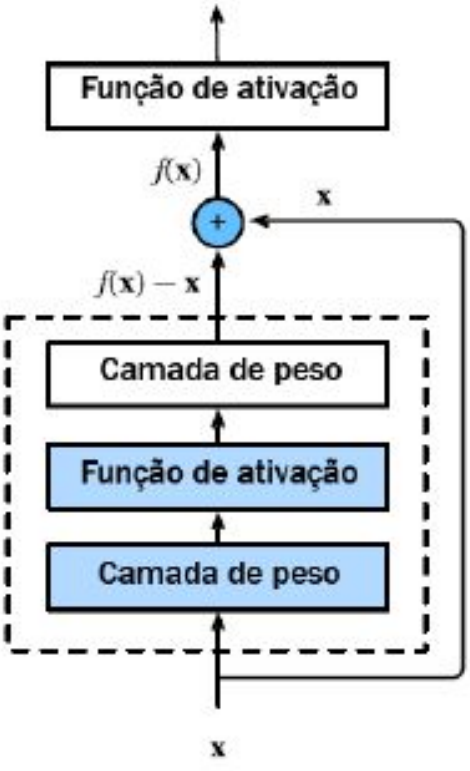


Figura 2.5 – Bloco residual de uma ResNet. Fonte: (FEITOSA, 2024)

Nessa arquitetura, a função residual F(x) é aprendida por meio do treinamento das camadas de peso com dados rotulados. Essas camadas podem ser de qualquer tipo comum em redes neurais, como camadas convolucionais ou totalmente conectadas (FC) (FEITOSA, 2024). O bloco residual permite que, durante a propagação direta, certas camadas sejam seletivamente ignoradas quando a função F(x) se aproxima de zero. Isso resulta em uma propagação mais eficiente, facilitada pelas conexões residuais entre as camadas.

Em redes muito profundas, um desafio conhecido é o problema do gradiente: à medida que a profundidade aumenta, os gradientes podem se tornar extremamente pequenos (vanishing) ou excessivamente grandes (exploding) durante a retropropagação, dificultando o ajuste dos pesos. As ResNets atenuam esse problema ao permitir que os gradientes fluam diretamente pelas conexões residuais, sem passar exclusivamente pelas camadas transformadoras. A relação fundamental que define o bloco residual é expressa

pela Equação 2.4:

$$y = F(x, W_i) + x \tag{2.4}$$

onde x representa a entrada do bloco residual, $F(x, \{W_i\})$ corresponde à função residual aprendida pelas camadas parametrizadas pelos pesos $\{W_i\}$, e y é a saída do bloco. De forma equivalente, pode-se interpretar que a rede busca aprender uma transformação residual definida por $F(x) = H(x) - x$, de modo que a função desejada seja reescrita como $H(x) = F(x) + x$ (FEITOSA, 2024).

É possível construir diferentes arquiteturas combinando diversos números de camadas e blocos residuais, como a *ResNet-18*, que possui 18 camadas no total. A *ResNet* exerceu grande influência em modelos subsequentes de redes neurais e continua sendo uma arquitetura de referência, tanto no contexto convolucional quanto em modelos sequenciais.

2.4.2 SqueezeNet

A *SqueezeNet* é uma arquitetura de rede neuronal convolucional (CNN) projetada com o objetivo principal de manter a precisão de modelos conhecidos, como a *AlexNet*, mas com um número drasticamente menor de parâmetros. Enquanto a *AlexNet* possui cerca de 240MB, a *SqueezeNet* consegue atingir o mesmo nível de precisão no dataset *ImageNet* com um tamanho de apenas 4,8MB (uma redução de 50x). Quando combinada com técnicas de compressão como a *Deep Compression*, o seu tamanho pode ser reduzido para menos de 0,5MB, tornando-a até 510x menor que a *AlexNet* sem perda de precisão (IANDOLA et al., 2016b).

Uma das vantagens fundamentais desta redução drástica no número de parâmetros é o ganho de eficiência em treino distribuído. Em arquiteturas *data-parallel*, a comunicação dos gradientes entre os servidores é frequentemente o principal gargalo de escalabilidade, e a sobrecarga dessa comunicação é diretamente proporcional ao número de parâmetros do modelo. Ao utilizar a *SqueezeNet*, a necessidade de comunicação entre servidores é substancialmente reduzida, permitindo que o modelo treine mais rapidamente. Além disso, o seu tamanho compacto facilita a atualização de modelos via rede (over-the-air) e a implementação em hardware com memória limitada, como FPGAs (IANDOLA et al., 2016b).

2.4.3 EfficientNet

As redes EfficientNet foram introduzidas por Tan e Le em 2019 (TAN; LE, 2019) como uma abordagem eficiente e sistemática para o escalonamento de arquiteturas de redes neurais convolucionais. Tradicionalmente, CNN são desenvolvidas com um orçamento fixo de recursos e, quando necessário, são ampliadas aumentando-se a profundidade, a largura ou a resolução das imagens de entrada de forma independente. Essa estratégia, entretanto, tende a resultar em modelos desbalanceados e pouco eficientes.

A EfficientNet propõe um método de *compound scaling*, no qual profundidade, largura e resolução são escalonadas de maneira conjunta e proporcional por meio de um único coeficiente composto. Esse balanceamento permite que o modelo preserve sua eficiência estrutural ao ser ampliado, resultando em ganhos significativos de desempenho.

Além disso, os autores empregaram a técnica de *Neural Architecture Search* (NAS) (TAN; LE, 2019) para definir uma arquitetura base, denominada *EfficientNet-B0*. A partir dessa configuração inicial, aplicaram um método de escalonamento composto para originar as variantes *EfficientNet-B1* até *B7*. Essa estratégia possibilitou um balanceamento sistemático entre profundidade, largura e resolução de entrada, promovendo um melhor compromisso entre acurácia e custo computacional. Como consequência, as *EfficientNets* superaram diversas arquiteturas clássicas, alcançando desempenho superior com um número significativamente menor de parâmetros e menor custo de inferência (TAN; LE, 2019).

2.4.4 Conv3D

As redes neurais convolucionais tridimensionais (*Conv3D*) surgiram como uma evolução natural das arquiteturas 2D para lidar com dados que possuem uma correlação intrínseca em uma terceira dimensão, como sequências temporais de vídeos ou volumes médicos (CT/MRI). Enquanto as convoluções tradicionais operam apenas nos eixos espacial de altura (h) e largura (w), a operação *Conv3D* aplica filtros que se deslocam simultaneamente em três direções. Isso permite que a rede extraia mapas de características espaciais e temporais de forma integrada, preservando a continuidade de informações que seria perdida se os dados fossem processados como fatias independentes (TRAN et al., 2015).

Do ponto de vista matemático, a principal distinção reside no formato do *kernel*

e na operação de convolução resultante. Em uma camada *Conv3D*, o filtro possui uma dimensão adicional denominada profundidade temporal (d), resultando em um volume de pesos. Ao deslizar esse cubo sobre o dado de entrada, o valor de cada neurônio na camada de saída é calculado através da soma do produto elemento a elemento entre o *kernel* e a sub-região correspondente do volume de entrada. Essa característica possibilita que o modelo capture padrões de movimento e mudanças de intensidade ao longo da sequência, tornando-a superior às abordagens 2D em tarefas de reconhecimento de ações e análise volumétrica (JI et al., 2012).

Diferente de modelos amplamente difundidos como a *ResNet* ou *EfficientNet*, que se beneficiam de pesos pré-treinados na base *ImageNet*, as redes *Conv3D* são frequentemente treinadas com inicialização aleatória dos parâmetros (*training from scratch*) em contextos específicos. Isso ocorre porque o pré-treinamento em larga escala para dados tridimensionais exige conjuntos de dados massivos e computacionalmente custosos, como o *Kinetics*. Assim, a arquitetura exige uma calibração rigorosa de hiperparâmetros e uma base de dados robusta o suficiente para que os filtros tridimensionais aprendam representações complexas sem o auxílio de transferência de aprendizado prévia (CARREIRA; ZISSERMAN, 2017).

2.5 Interpretabilidade em Inteligência Artificial

A interpretabilidade é um conceito que vem ganhando atenção no campo da IA. Ele se refere à capacidade de compreender e explicar o funcionamento e as decisões de um modelo de *machine learning* de maneira clara e intuitiva (COELHO, 2024). Em outras palavras, trata-se da capacidade de um sistema de IA de ser compreendio e justivicado por seres humanos, especialmente por especialistas não técnicos, como profissionais de diversas áreas que podem não ter conhecimento avançado em ciências de dados ou aprendizado de máquina. Diversas técnicas são empregadas para alcançar a interpretabilidade, incluindo a visualização de dados, a simplificação de modelos complexo, a geração de explicações compreensíves e a identificação de padrões relevantes nos processos de tomada de desisão (COELHO, 2024). Essas técnicas são fundamentais para promover a confiança e a transparência na aplicação de sistemas de IA. A compreensão da interpretabilidade de modelos de aprendizado de máquina abrange uma variedade de abordagens, sendo notáveis os métodos que se baseiam na análise das ativações em redes neurais convolucinais predefinidas. Esses métodos buscam examinar o comportamento de camadas convolucio-

nais, mapeando as projeções dos níveis de ativação resultantes (COELHO, 2024).

2.5.1 Grad-CAM

Na tentativa de compreender melhor as CNNs, vários métodos surgia na literatura para visualização da representações interna desse tipo de rede. Uma das abordagens criadas foram os *Class Activation Map* (CAM), cujo objetivo é utilizar *Global Average Pooling* (GAP) nas camandas da CNN (FEITOSA, 2024). Assim, esse método é baseado na utilização da GAP no mapa de ativação da última camada convolucional da rede, ante da FC. Dessa forma, CAM combina os mapas de ativação A da camada de convolução, que contém, que contém *K* filtros, alem dos pesos $w_{k,c}$ de FC, cujos valores (k, c) representam a consexão ponderada específica entre a camada de convolução e a camada totalmente conectada utilizada para criar a pontuação do mapa de relevância, conforme a Equação 2.5.

$$\text{map}_c = \sum_{k}^{K} w_{k,c}\, A_k \tag{2.5}$$

O método Grad-CAM é uma variação do método CAM usando os gradientes da saída da rede em relação à última camada convolucional da CNN, a fim de obter o mapa de ativação de classe (FEITOSA, 2024). Com isso, é possível que o Grad-CAM seja implementado em mais tipos de CNNs em comparação ao método CAM, exigindo apenas que a função de ativação final usada para previsão de rede seja uma função diferenciável, como por exemplo, a função *softmax*. Dessa forma, para cada mapa de características na camada de convolução final, um gradiente da pontuação y_c *(logit)* da classe em relação a todo nó em A_k é calculado para obter uma pontuação de importancia $\alpha_{k,c}$ para o mapa de características A_k. A partir disso, o método Grad-CAM combina linearmente essas pontuações de importância e as passa pela função ReLU para obter um mapa de relevância, conforme a Equação 2.6.

$$\text{map}_c = ReLU(\sum_{k}^{K} \alpha_{k,c} A_k) \tag{2.6}$$

A principal diferença entre CAM e Grad-CAM está no modo como os pesos utilizados para gerar os mapas de características são obtidos. No CAM, os mapas de calor são produzidos por uma soma ponderada das ativações da última camada convolucional,

utilizando diretamente os pesos da camada totalmente conectada (FC). Já no Grad-CAM, esses pesos não dependem da estrutura da FC: eles são calculados a partir dos gradientes da classe de interesse em relação aos mapas de ativação de qualquer camada convolucional escolhida (FEITOSA, 2024). Essa distinção é ilustrada na Figura 2.6, onde ambos os métodos foram aplicados à classe “pessoa” em uma mesma imagem para fins de comparação.

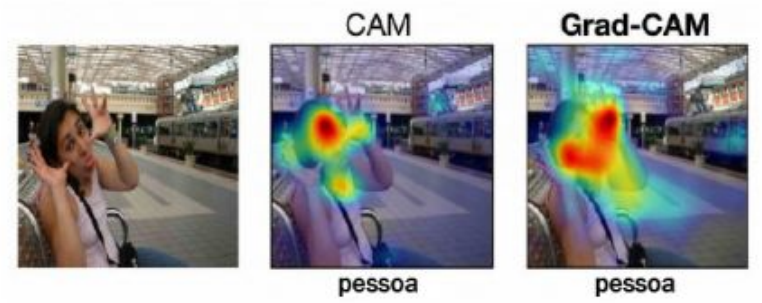


Figura 2.6 – Fonte (FEITOSA, 2024)

Como visto na Fig. 2.6, os gradientes globais são menos precisos e operam sobre mapas de ativação de baixa resolução, o que resulta em regiões de destaque mais amplas e menos localizadas.

2.5.1.1 Layer Grad-CAM

Quando o método Grad-CAM é aplicado a uma determinada camada, essa obordagem é chamada de *Layer Grad-CAM*Layer Grad-CAM (FEITOSA, 2024). Portanto os gradientes da saída são calculados em relação à camada escolhida cujos gradientes resultantes são calculados em média para cada canal de saída. Assim, o gradiente médio de cada canal é multiplicado pelas ativações da camada e em seguida, os resultados são somados em todos os canais. pra os pesos w referente a característica da classe c, o GAP é realizado sobre os mapas de recursos A, de acordo com a Equação 2.7

$$S^C = \sum_k w_k^c \frac{1}{Z} \sum_i \sum_j A_{i,j}^k \tag{2.7}$$

2.5.2 Mapa de Saliência

Mapas de Saliência ou Mapas de Calor é um método de visualização comumente utlizado para explicar o processo de predição de uma rede. Dessa forma, muitas das

vezes que um mapa de calor é sobreposto à imagem de entrada original para que assim seja gerada a visualiação. Além disso, esses mapas identificam caraterísticas da entra que são mais relevantes (salientes). Ou seja, destaca regiões que estimula o máximo a rede, de forma a influenciar a saída do modelo (FEITOSA, 2024). A extração da saliência da classe ocorre dada uma imagem I_0 (com m linhas e n colunas) e uma classe c. Assim, o Mapa de Saliência da classe M é calculado a partir da derivada w que é encontrada por retropropagação. Depois disso, o Mapa de Saliência é obtido reorganizando os elementos do vetor w. Caso a imagem esteja em escala de cinza, o número de elemtnso em w é igual ao número de pixels em I_0, possibilitando que o mapa seja calculado como na Equação 2.8, cujo valor $h(i, j)$ equivale ao índice do elemento de w correpondente ao pixel da imagem na *i*-ésima linha e *j*-ésima coluna.

$$M_{i,j} = |w_{h(i,j)}| \tag{2.8}$$

Segundo (FEITOSA, 2024), no caso da imagem colorida, assume-se, por exemplo, que o canal de cor c do pixel (i, j) da imagem I corresponde ao elemento de w com o índice $h(i, j, c)$. Dessa forma, para derivar um valor de saliência de classe exclusivo para cada pixel (i, j), é necessário selecionar a magnitude máxima de w em todos os canais de cores, conforme a Equação 2.9.

$$M_{i,j} = max_c|w_{h(i,j,c)}| \tag{2.9}$$

3 Trabalhos relacionados

Este capítulo apresenta a metodologia adotada para a realização do levantamento bibliográfico, fundamentada nos princípios da Revisão Sistemática da Literatura (RSL). São descritas as etapas de planejamento, condução e documentação da revisão, bem como a definição das questões de pesquisa estruturadas a partir do modelo PICOC. Além disso, detalham-se os critérios de busca e exclusão utilizados na seleção dos estudos, com o objetivo de garantir rigor metodológico, transparência e confiabilidade na análise do estado da arte acerca das técnicas de explicabilidade aplicadas à visão computacional em imagens médicas.

3.1 Metodologia do Levantamento Bibliográfico

Neste capítulo, são apresentados os resultados obtidos com base em elementos característicos da Revisão Sistemática da Literatura (RSL) em relação às questões de pesquisa previamente definidas.

A utilização desses elementos proporciona vantagens significativas para este estudo, como a obtenção de resultados mais consistentes, evidências mais confiáveis e uma compreensão mais aprofundada do estado da arte sobre o tema investigado (FEITOSA, 2024). Com base nisso, descrevem-se a seguir as etapas que geralmente compõem uma RSL — planejamento, condução da revisão e documentação — que auxiliam no processo de identificação, seleção e avaliação de documentos.

Na etapa de planejamento, as questões de pesquisa foram estruturadas utilizando o modelo PICOC (População, Intervenção, Comparação, Saídas e Contexto), com o objetivo de organizar os termos fundamentais que orientam a revisão.

Assim, os parâmetros do modelo adotados neste trabalho foram definidos da seguinte forma:

- **População:** imagens médicas, especialmente exames de ressonância magnética (MRI);
- **Intervenção:** técnicas, métodos e ferramentas de explicabilidade aplicadas a modelos de visão computacional;
- **Comparação:** diferentes métodos, modelos ou abordagens utilizados na geração de explicabilidade;
- **Saídas:** evidências, métricas ou resultados relacionados à explicabilidade e interpretabilidade;
- **Contexto:** estudos e aplicações inseridos na área de visão computacional.

Todos os parâmetros foram utilizados em inglês durante o levantamento bibliográfico, a fim de maximizar a recuperação de documentos relevantes, uma vez que a maior parte das publicações está nessa língua. Com base no PICOC definido, foram elaboradas questões de pesquisa que orientam este trabalho:

- *Q1: Quais métodos de explicabilidade têm sido utilizados para interpretar modelos de visão computacional aplicados a imagens médicas?*
- *Q2: Quais arquiteturas de redes neurais (CNN, EfficientNet, ResNet, entre outras) são mais frequentemente combinadas com técnicas de explicabilidade no contexto da detecção de Alzheimer?*
- *Q3: Quais limitações têm sido apontadas na literatura sobre o uso de técnicas de explicabilidade em imagens médicas, especialmente no diagnóstico assistido por IA?*

Durante a condução da pesquisa bibliográfica, essas questões serviram como referência para compreender o estado atual das técnicas de explicabilidade aplicadas à visão computacional e para identificar métodos existentes que aprimoram a transparência e a confiabilidade das explicações geradas. Para a busca de publicações, foram consideradas bibliotecas como *ScienceDirect* e *Google Scholar*, e aplicados critérios de exclusão como:

- Documentos publicados antes de 2021;
- Tutoriais, resumos de eventos, pôsteres ou outros materiais incompletos;
- Títulos, resumos ou palavras-chave sem relação com o tema investigado;
- Documentos duplicados;
- Documentos indisponíveis para *download*.

3.2 Resultados

Os resultados serão apresentados de modo a responder a cada uma das perguntas de pesquisa.

3.2.1 Q1: Métodos de explicabilidade em modelos de visão computacional aplicados a imagens médicas

Os métodos mais citados na literatura incluem abordagens baseadas em *Class Activation Maps* (CAM) e suas extensões, como o *Grad-CAM* (Seção 2.5.1), que fornecem mapas de calor indicando as regiões da imagem mais relevantes para a predição do modelo.

Além disso, técnicas independentes de arquitetura, como *SHAP* (LUNDBERG; LEE, 2017) e *LIME* (RIBEIRO; SINGH; GUESTRIN, 2016), têm sido aplicadas de forma complementar. O SHAP quantifica a contribuição de cada característica para a decisão do modelo, enquanto o LIME gera explicações locais por meio de perturbações na entrada.

Alguns estudos utilizam *Decision Trees* (DT) como *modelos substitutos locais* (*local surrogate models*) para aproximar o comportamento de modelos complexos, permitindo explicações mais interpretáveis (SANTOS et al., 2022).

De forma geral, observa-se a tendência de combinar métodos específicos para imagens com técnicas independentes do modelo, oferecendo análises mais robustas e completas.

3.2.2 Q2: Arquiteturas de redes neurais mais utilizadas com técnicas de explicabilidade na detecção de Alzheimer

As arquiteturas mais frequentes em estudos relacionados à detecção ou classificação da Doença de Alzheimer incluem *EfficientNet* (Seção 2.4.3), *ResNet* (Seção 2.4.1), *DenseNet* (HUANG et al., 2017; HASSAN et al., 2025), *Inception* (SZEGEDY et al., 2016) e *MobileNet* (HOWARD et al., 2017). Modelos específicos mencionados incluem *EfficientNet-B0/B2*, *ResNet50/101*, *DenseNet121*, *Inception V3*, *MobileNet* e *MobileNetV3*, evidenciando a preferência por arquiteturas que equilibram desempenho e custo computacional. Modelos clássicos como a *VGG-19* (SIMONYAN; ZISSERMAN, 2014) apare-

cem com menor frequência.

CNNs continuam sendo a principal escolha quando combinadas com técnicas de explicabilidade visual, permitindo identificar regiões relevantes das imagens e oferecer maior transparência às decisões do modelo.

3.2.3 Q3: Limitações das técnicas de explicabilidade em imagens médicas

A aplicação da explicabilidade clínica enfrenta obstáculos que transcendem a precisão dos modelos. O desbalanceamento de classes em bases como ADNI (JR et al., 2008) e OASIS-1 (MARCUS et al., 2007) é um desafio central, pois pode induzir interpretações baseadas em atalhos estatísticos em vez de biomarcadores reais. Além disso, a natureza do dado — seja transversal (cross-sectional), como no OASIS-1, ou de acompanhamento (longitudinal), como no OASIS-3 — condiciona o alcance da análise: enquanto dados pontuais oferecem visões estáticas, os longitudinais exigem que a explicabilidade contemple a progressão temporal da doença. Outro fator crítico é a perda de contexto espacial ao utilizar cortes bidimensionais em exames volumétricos, resultando na negligência de informações como a atrofia completa do hipocampo (WEN et al., 2020a), o que descaracteriza a continuidade anatômica. Essas limitações técnicas culminam em dilemas éticos e regulatórios (ALABI, 2025), evidenciando que a transparência na saúde exige um rigor metodológico que considere as particularidades intrínsecas dos dados médicos.

3.3 Resumo

Observou-se uma escassez de trabalhos introdutórios que apresentem, de forma acessível, os desafios iniciais do uso de técnicas de explicabilidade em modelos de aprendizado de máquina aplicados à saúde. A maioria das pesquisas concentra-se em abordagens avançadas voltadas a modelos complexos, como evidenciado em (SANTOS et al., 2022), que utiliza uma árvore de decisão como estrutura explicativa simples. Esse panorama evidencia a necessidade de estudos mais didáticos, que abordem conceitos fundamentais, dificuldades práticas e limitações comuns — tais como dados desbalanceados e a seleção de imagens relevantes —, facilitando a compreensão e a adoção de técnicas de explicabilidade em contextos clínicos sensíveis.

4 Metodologia

Este capítulo apresenta as arquiteturas de redes neurais convolucionais utilizadas na etapa experimental do trabalho. São abordados modelos pré-treinados representativos de diferentes gerações de CNNs, bem como a arquitetura convolucional tridimensional (Conv3D), adequada para dados volumétricos. O objetivo é descrever as principais características dessas abordagens e justificar sua aplicação na tarefa de classificação de imagens médicas proposta.

4.1 Modelos de CNN Pré-Treinados

Neste capítulo, apresentamos as quatro arquiteturas de modelos pré-treinados que serão adotadas na etapa experimental deste trabalho: *ResNet18*, *AlexNet*, *SqueezeNet* e *EfficientNet-B3*.

Essas arquiteturas selecionatas representam diferentes gerações de CNNs:

- modelo clássico (*AlexNet*),
- modelo compacto (*SqueezeNet*),
- modelo com conexões residuais (*ResNet-18*) e
- modelo otimizado por escalonamento composto (*EfficientNet-B3*).

A rede *AlexNet* é considerada um marco no desenvolvimento do *Deep Learning*, tendo vencido a competição ImageNet em 2012. Sua arquitetura é composta por oito camadas de aprendizado, sendo cinco camadas convolucionais seguidas por três camadas totalmente conectadas (*fully connected*). A *AlexNet* popularizou o uso de funções de ativação *ReLU* (Rectified Linear Unit) para acelerar o treinamento e a técnica de *Dropout* para reduzir o *overfitting*, estabelecendo as bases para as CNNs modernas (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

Por outro lado, a *SqueezeNet* foi desenvolvida com o objetivo específico de criar

uma arquitetura extremamente leve, mantendo uma acurácia comparável à da AlexNet, mas com 50 vezes menos parâmetros. O projeto baseia-se no módulo *Fire*, que consiste em uma camada de compressão (*squeeze*) composta por filtros 1×1, seguida por uma camada de expansão (*expand*) com uma combinação de filtros 1×1 e 3×3. Essa estratégia minimiza o esforço computacional e o tamanho do modelo, facilitando sua implementação em dispositivos com recursos limitados (IANDOLA et al., 2016a).

A rede *ResNet18* faz parte de uma família de redes convolucionais denominada *Residual Network*. Elas foram projetadas para superar desafios no treinamento de redes muito profundas, introduzindo o conceito de aprendizado residual (*residual learning*) (CASTRO et al., 2025).

A *EfficientNet-B3* foi projetada para alcançar um equilíbrio ideal entre acurácia e eficiência computacional através de um método de escalonamento composto (*compound scaling*). Diferente de abordagens convencionais, essa arquitetura escala simultaneamente a profundidade, a largura e a resolução da imagem de entrada utilizando um coeficiente composto uniforme. A estrutura base fundamenta-se em blocos de convolução invertida residual (*MBConv*), incorporando mecanismos de *Squeeze-and-Excitation* para recalibrar as respostas dos canais e otimizar o fluxo de características, o que permite manter um alto desempenho na classificação de imagens com um número reduzido de parâmetros (TAN; LE, 2019).

4.2 Arquitetura Convolucional Tridimensional (Conv3D)

Diferente das arquiteturas bidimensionais previamente descritas, a camada *Conv3D* é projetada para processar dados volumétricos ou sequências temporais, como vídeos ou exames médicos volumétricos. Enquanto as convoluções 2D operam apenas nas dimensões de altura e largura, a convolução 3D estende esse processo para uma terceira dimensão, permitindo a extração de características espaciais e temporais simultaneamente através de um *kernel* tridimensional. Esta capacidade é fundamental para capturar o movimento e a correlação entre *frames* consecutivos (TRAN et al., 2015).

Dada a natureza específica dos dados de entrada, este modelo foi utilizado sem pesos pré-treinados, sendo submetido a um processo de inicialização aleatória dos parâmetros. A ausência de pré-treinamento em bases como a *ImageNet* — que é composta predominantemente por imagens estáticas 2D — justifica-se pela incompatibilidade estrutural dos filtros, exigindo que a rede aprenda as representações hierárquicas do zero

(*from scratch*) diretamente a partir do conjunto de dados experimental deste trabalho (JI et al., 2012).

4.2.1 Redes Neurais Utilizadas

Inicialmente, a abordagem prevista consistia na utilização de redes convolucionais (CNN) e Transformers para análise das imagens do dataset. Considerando que o conjunto de dados continha aproximadamente 80.000 imagens, optou-se por modelos que equilibrassem **desempenho e eficiência computacional**, permitindo treinamento em ambientes gratuitos como Kaggle e Google Colaboratory.

As arquiteturas selecionadas para os experimentos foram: **SqueezeNet**, **ResNet18**, **EfficientNet-B3** e **AlexNet**. Todas se mostraram adequadas para o contexto do projeto, oferecendo um bom compromisso entre complexidade computacional e capacidade de extração de características das imagens.

4.2.2 Dataset Utilizado

Para este trabalho, utilizou-se o dataset **OASIS Alzheimer's Detection**, que consiste em um conjunto de imagens de ressonância magnética cerebral (MRI) voltado para análise com redes neurais profundas. Os dados originais foram obtidos via plataforma Kaggle, disponível em: <https://www.kaggle.com/datasets/ninadaithal/imagesoasis>.

O dataset original contém aproximadamente 80.000 imagens de ressonância magnética cerebral. As imagens foram divididas em quatro classes, com base na progressão da doença de Alzheimer, de acordo com o *Clinical Dementia Rating (CDR)*:

- *Non-demented* (sem demência)
- *Very mild demented* (demência muito leve)
- *Mild demented* (demência leve)
- *Moderate Dementia* (demência moderada)

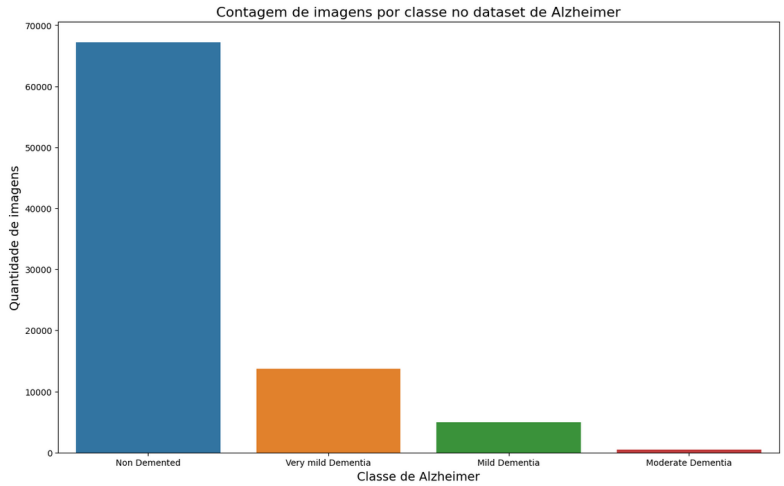


Figura 4.1 – Distribuição da quantidade de imagens por classe diagnóstica no dataset OASIS-1.

Como apresentado na Figura 4.1, o dataset OASIS-1 apresenta um desbalanceamento acentuado, sendo composto majoritariamente por dados da classe *Non Demented*, o que reflete a distribuição comum em bases de dados de rastreio clínico.

Os arquivos originais, inicialmente nos formatos `.img` e `.hdr`, foram convertidos para o padrão NIfTI (`.nii`) utilizando a biblioteca FSL (*FMRIB Software Library*). Para viabilizar o processamento por meio de redes neurais 2D, as imagens foram fatiadas ao longo do eixo axial (transversal), conforme ilustrado na Figura 4.2.

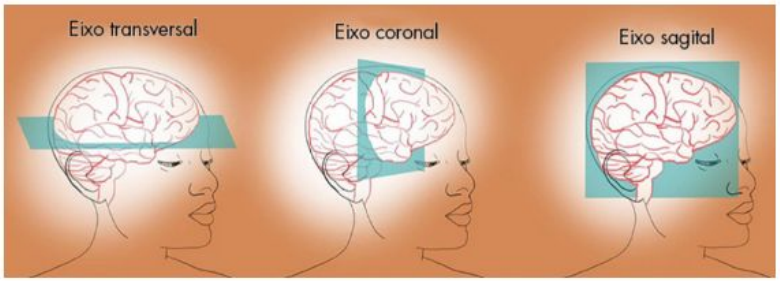


Figura 4.2 – Eixos de análise retirado de (FILHO; VIDAL, 2024).

As imagens estavam no formato `.jpeg`, o que impossibilitou a separação por eixos anatômicos. Assim, a única forma de organização viável foi a divisão das imagens por paciente, totalizando 256 fatias por indivíduo. Em contraste, outros trabalhos (FILHO; VIDAL, 2024) utilizam o formato *NIfTI*, que permite a manipulação e separação da base de dados por eixos, como ilustrado na Figura 4.2.

Esta base de dados permite explorar diferentes modelos de redes neurais na detecção precoce e análise das diversas fases da doença de Alzheimer.

Este trabalho adota uma abordagem experimental utilizada em estudos anteriores (GOMIDES, 2024; SOUSA et al., 2025; RODRIGUES, 2024) com o objetivo de compa-

rar o desempenho de arquiteturas de redes neurais convolucionais (CNNs) e de modelos baseados em transformers na tarefa de classificação de imagens médicas. A metodologia empregada contempla as etapas de preparação dos dados, desenvolvimento e treinamento dos modelos, avaliação quantitativa dos resultados e análise comparativa das métricas obtidas.

4.2.3 Divisão do Conjunto de Dados por Paciente

A organização do conjunto de dados foi realizada com base no identificador (ID) de cada paciente, garantindo que todas as imagens de um mesmo indivíduo pertencessem exclusivamente a um único grupo (treino, validação ou teste). Esta etapa é crítica para evitar o *data leakage* (vazamento de dados), fenômeno que ocorre quando informações do conjunto de teste influenciam o treinamento, resultando em métricas de desempenho artificialmente infladas e modelos com baixa capacidade de generalização.

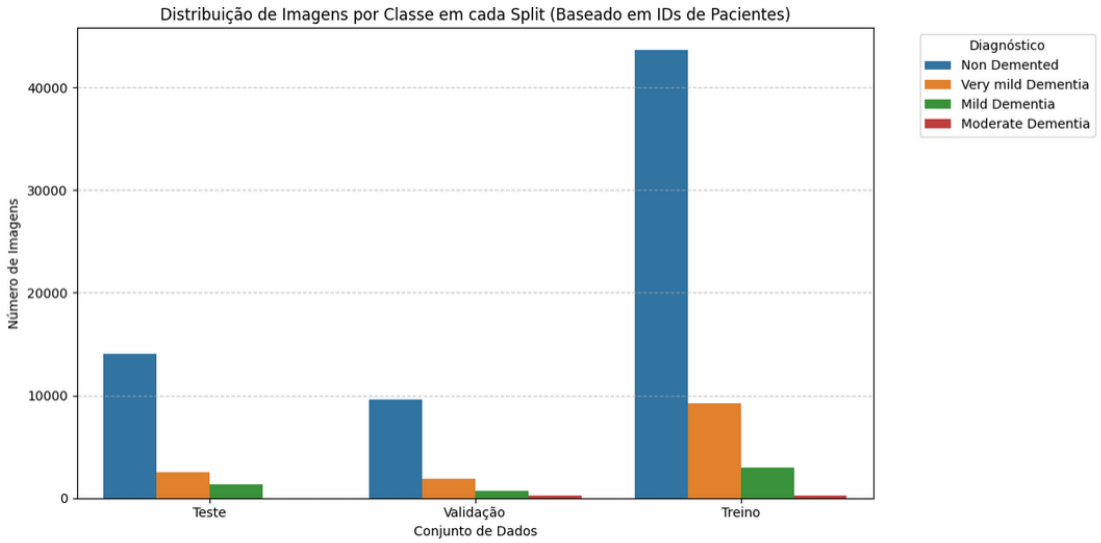


Figura 4.3 – Distribuição dos dados por paciente e separação em treino, teste e validação.

A base de dados utilizada compreende um total de 347 pacientes. Para o desenvolvimento dos modelos, os dados foram particionados (Figura 4.3) da seguinte forma:

- **Treino:** 225 pacientes (aprox. 65%), destinados ao aprendizado dos pesos das redes neurais;
- **Validação:** 52 pacientes (aprox. 15%), utilizados para o ajuste de hiperparâmetros e monitoramento do *overfitting*;
- **Teste:** 70 pacientes (aprox. 20%), reservados para a avaliação final e imparcial da

capacidade de generalização das arquiteturas propostas.

Diferente de diversas abordagens na literatura que realizam a divisão de forma aleatória por imagens, negligenciando a dependência entre múltiplos cortes diagnósticos do mesmo paciente (WEN et al., 2020b), esta metodologia assegura a robustez estatística necessária para aplicações em diagnóstico auxiliado por computador.

Para garantir a integridade da validação do modelo, implementou-se uma estratégia de divisão dos dados estritamente baseada em pacientes, evitando o vazimento de dados entre os conjuntos de treino, validação e teste. Aplicou-se oversampling na classe Mild Dementia com fator de $2\times$, aumentando sua representação para 10.002 imagens. A divisão estratificada resultou em 59.387 imagens para treino, 13.705 para validação e 18.274 para teste, mantendo a proporção original das classes em todos os subconjuntos.

A alocação de uma proporção maior para o conjunto de teste (20%) em detrimento da validação (15%) fundamenta-se na necessidade de garantir a estabilidade das métricas de desempenho. No domínio médico, o conjunto de teste deve ser suficientemente robusto para mitigar a variância na estimativa do erro de generalização, assegurando que os resultados reportados não sejam submetidos a vieses de amostragem (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Dessa forma, priorizou-se um n amostral mais elevado na avaliação final para permitir uma análise de erro mais fidedigna e representativa da diversidade clínica, uma prática recomendada para validar a confiabilidade de sistemas de diagnóstico auxiliado por computador (CAD) antes de sua transposição para o cenário clínico real (VAROQUAUX; COLLIOT, 2023).

4.3 Configurações Experimentais

Para todos os experimentos realizados neste trabalho, adotou-se a seguinte configuração de treinamento da rede neural:

As configurações de treinamento foram definidas buscando equilibrar custo computacional e qualidade do processo de aprendizado da rede neural.

Para garantir uma comparação justa entre os modelos avaliados, foi adotado um conjunto padronizado de hiperparâmetros, sem realizar otimizações específicas para cada arquitetura. O objetivo foi analisar o desempenho relativo das redes sob as mesmas condições experimentais. Assim, foram utilizados valores comuns na literatura para tarefas

Tabela 4.1 – Hiperparâmetros utilizados nos experimentos.

Parâmetro	**Valor**
Dispositivo de execução	`cuda` (GPU) / `cpu`
Tamanho do batch (Batch size)	32
Taxa de aprendizado (Learning rate)	1×10^{-4}
Número de épocas (Epochs)	50
Step size do scheduler	5
Fator de decaimento (Gamma)	0.1
Número de workers	2
Pin memory	True
Semente aleatória (Random seed)	42

de classificação de imagens, como taxa de aprendizado de 1×10^{-4}, *batch size* de 32 e treinamento com até 50 épocas. Além disso, foi aplicado o mecanismo de *early stopping*, permitindo interromper o treinamento quando não há melhoria no desempenho do modelo, reduzindo o risco de *overfitting*.

Apesar da premissa de manter o conjunto de dados original, foi necessário aplicar uma técnica de aumento de dados (*data augmentation*) pontual para as classes minoritárias. Essa medida foi adotada especificamente para viabilizar o treinamento da arquitetura *AlexNet*, que apresentava instabilidade devido ao desbalanceamento severo: a ausência de amostras de determinadas classes em alguns batches impedia a convergência do modelo. Assim, utilizou-se o sobrebalanceamento (*oversampling*) por meio de transformações geométricas simples, garantindo a representatividade de todas as categorias em cada iteração e a continuidade do processo de aprendizado.

4.4 Grad-CAM

A técnica Grad-CAM (Gradient-weighted Class Activation Mapping) foi selecionada para a interpretação dos modelos devido à sua eficiência e baixo custo computacional. Essa característica permitiu a geração de mapas de relevância em ambientes com recursos de processamento limitados, como o *Google Colab* e o *Kaggle*. A implementação foi realizada de forma manual utilizando a biblioteca *PyTorch* para a extração de gradientes e ativações, o *NumPy* para o processamento numérico, e as bibliotecas *OpenCV* e *Matplotlib* para a renderização e visualização dos mapas de calor sobre as imagens originais.

5 Resultados

Este capítulo apresenta os resultados obtidos a partir da execução dos modelos de redes neurais convolucionais 2D e da arquitetura *Conv3D* aplicados à classificação de estágios da Doença de Alzheimer. São analisadas métricas quantitativas de desempenho, matrizes de confusão e mapas de ativação Grad-CAM, com o objetivo de avaliar a capacidade discriminativa, a robustez frente ao desbalanceamento dos dados e a influência de fatores como resolução das imagens e estratégia de treinamento. A comparação entre as abordagens permite discutir vantagens, limitações e implicações práticas para aplicações em contexto clínico.

5.1 Resultados da Execução dos Modelos CNN 2D

Uma das principais vantagens do modelo *SqueezeNet* é o tamanho reduzido do arquivo do modelo, que possui apenas 8.4 MB, facilitando a implantação em dispositivos com recursos limitados (edge devices).

Tabela 5.1 – Comparativo de métricas de desempenho: *AlexNet*, *SqueezeNet* , *ResNet-18* e *EfficientNet-B3*.

Métrica	***AlexNet***	***SqueezeNet***	***ResNet-18***	***EfficientNet-B3***
Acurácia	79,81 %	80,16 %	99,95 %	66,16 %
Precisão	50,87 %	48,83 %	99,93 %	75,33 %
Recall	52,01 %	51,43 %	99,95 %	66,16 %
F1-Score	51,34 %	49,37 %	99,94 %	69,69 %
Época	18	25	29	45

Também foi realizada uma investigação sobre como o tamanho das imagens pode influenciar o desempenho dos modelos. Para isso, o modelo foi avaliado utilizando imagens nas resoluções 348 e 300, e os resultados obtidos estão apresentados na Tabela 5.2,

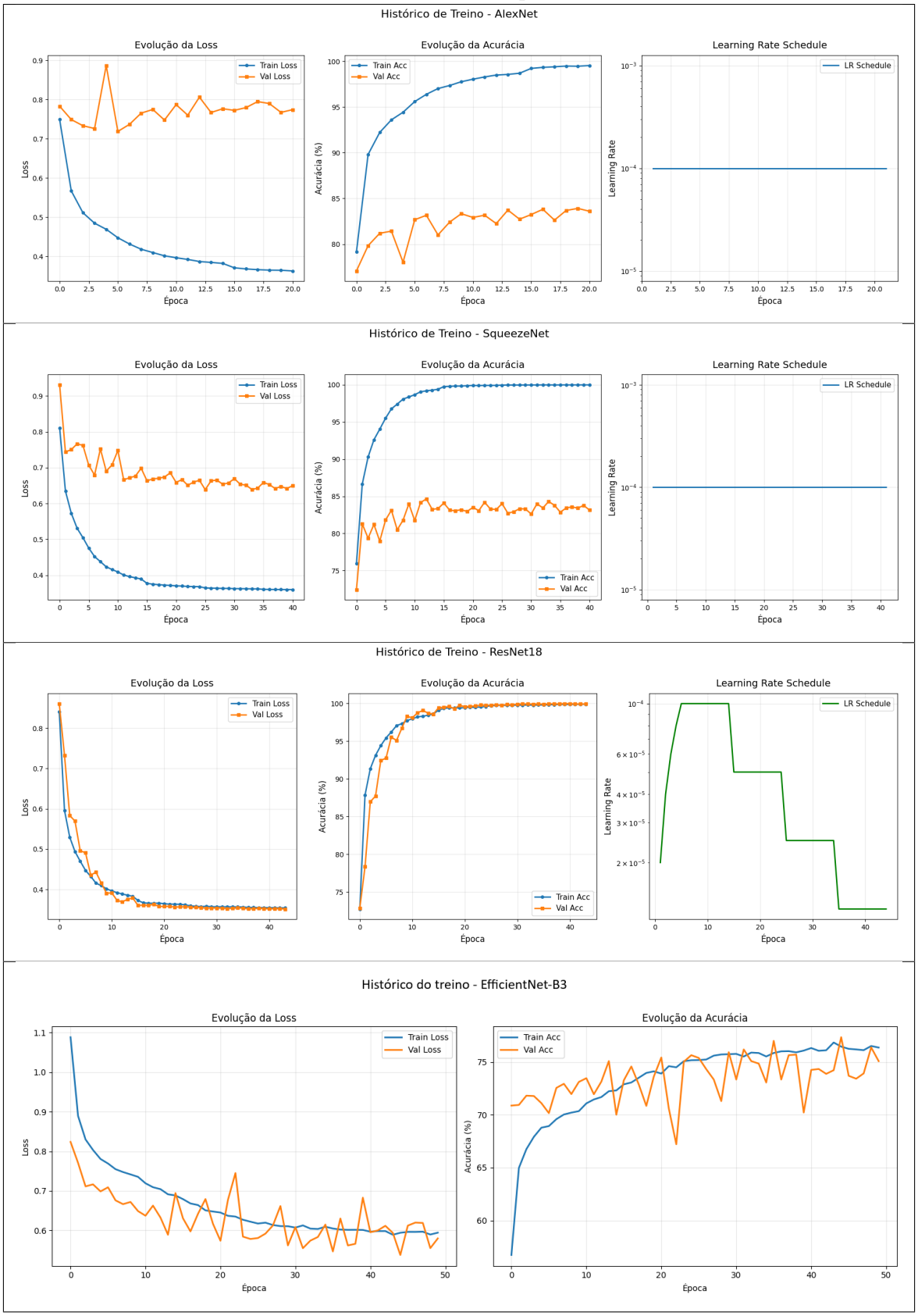
esses mostram que, embora a resolução *348×348* proporcione um tempo de treinamento por época significativamente menor, o desempenho global do modelo é inferior em todos os indicadores. A resolução *300×300*, utilizada na *EfficientNet-B3*, apresentou métricas superiores — com aumento de aproximadamente 3–4% na acurácia, precisão, *recall* e *F1-score*. Esses resultados sugerem que a resolução *300×300* é mais adequada para a arquitetura *EfficientNet-B3*, proporcionando melhor qualidade discriminativa mesmo com maior custo computacional por época.

Tabela 5.2 – Comparativo de métricas de desempenho: *EfficientNet-B3* em diferentes resoluções.

Métrica	**EfficientNet-B3 (300 $\times$ 300)**	**EfficientNet-B3 (348 $\times$ 348)**
Acurácia	65,18 %	61,68 %
Precisão	76,08 %	74,30 %
Recall	65,18 %	61,68 %
F1-Score	69,10 %	66,47 %
Época	36	18

Segundo a documentação oficial do *Pytorch* (TORCHVISION...,) o tamanho correto do modelo e 300×300.

Tabela 5.3 – Histórico de treino dos modelos *AlexNet*, *SqueezeNet*, *ResNet-18* e *EfficientNet-B3*.



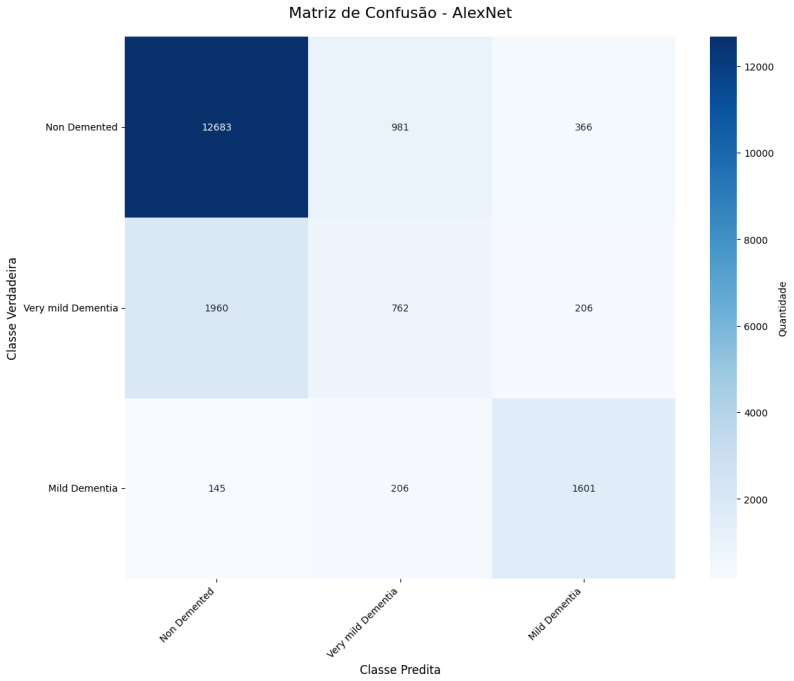


Figura 5.1 – Matriz de Confusão *AlexNet*

Como os dados estavão muito desbalanceado a classe minoritária *Moderate Dementia* ficou com poucos pacientes no momento da divisão.

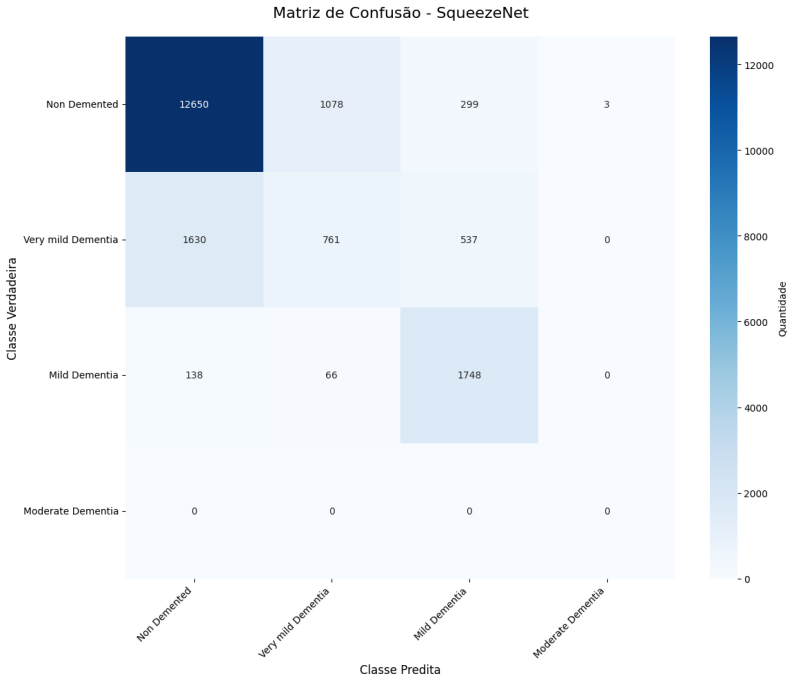


Figura 5.2 – Matriz de Confusão *SqueezeNet*

Apesar da alta acurácia (esperada para este dataset devido à grande similaridade entre as imagens), o modelo final ainda apresenta um tamanho considerável de 44.7 MB.

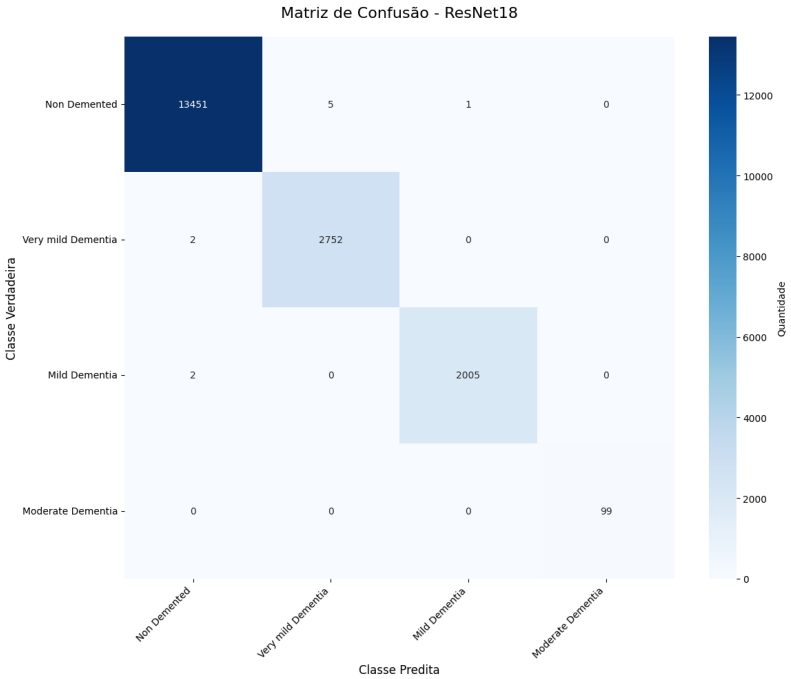


Figura 5.3 – Matriz de confusão *ResNet18*

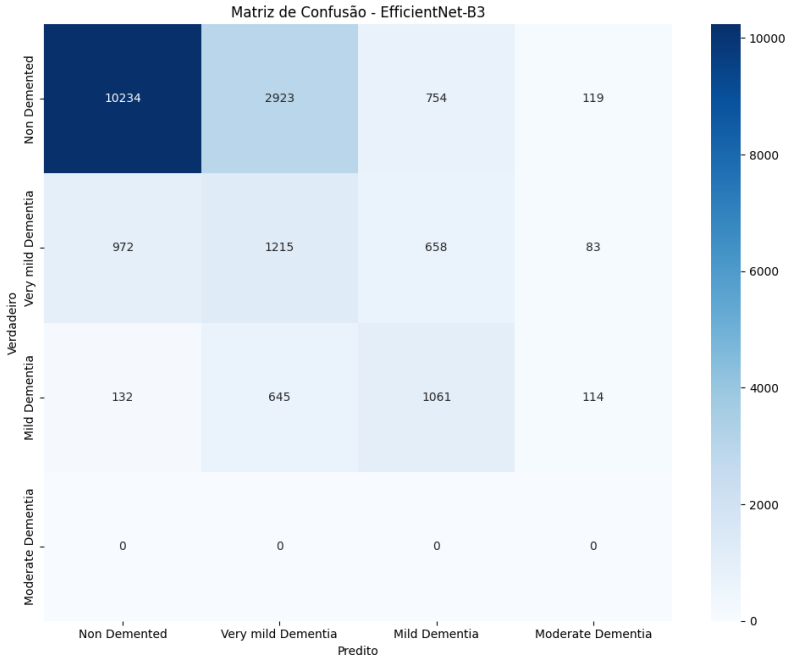


Figura 5.4 – Matriz de confusão *EfficienteNet-B3*

As Figuras 5.5, 5.6 e 5.7 mostram os mapas Grad-CAM dos modelos AlexNet, SqueezeNet e ResNet18, respectivamente. Em cada figura, são apresentados o *slice* analisado, o mapa de calor gerado e a imagem resultante da sobreposição do mapa à imagem original.

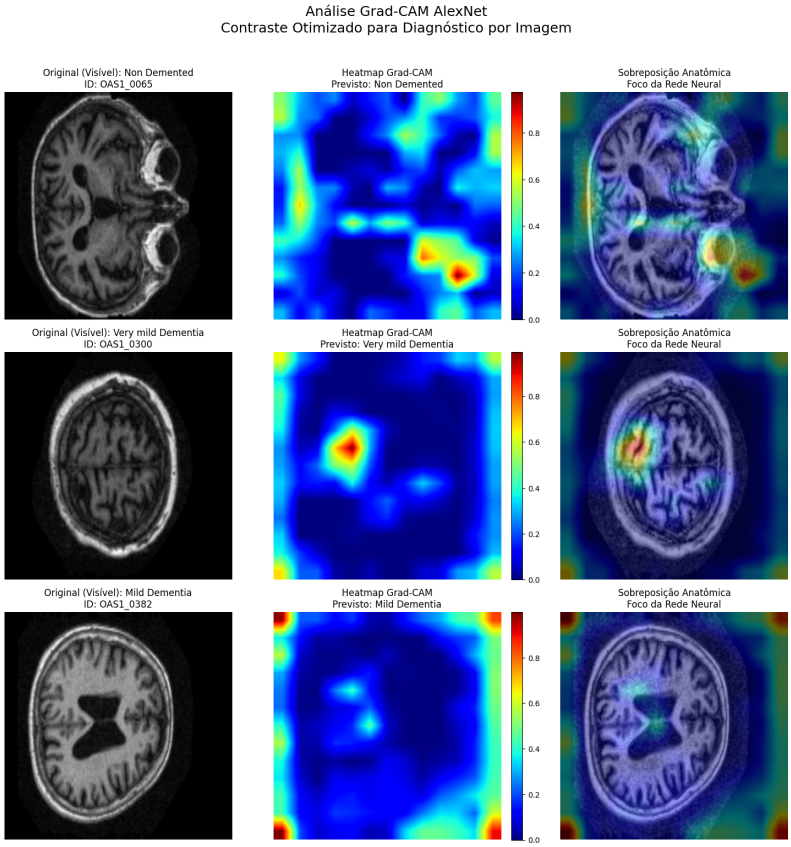


Figura 5.5 – Mapa Grad-CAM do modelo *AlexNet*

Identificou-se, por meio do Grad-CAM (Figura 5.5), que o modelo prioriza artefatos posicionais periféricos em detrimento das nuances patológicas centrais. Conforme sugerido por (SHORTEN; KHOSHGOFTAAR, 2019), a aplicação de *Random Resized Crop* é uma estratégia eficaz para mitigar esse sobreajuste (*overfitting*). Ao variar as subregiões analisadas durante o treinamento, a técnica compele o algoritmo a extrair padrões mais robustos do parênquima cerebral, otimizando a capacidade de generalização e a confiabilidade do modelo final.

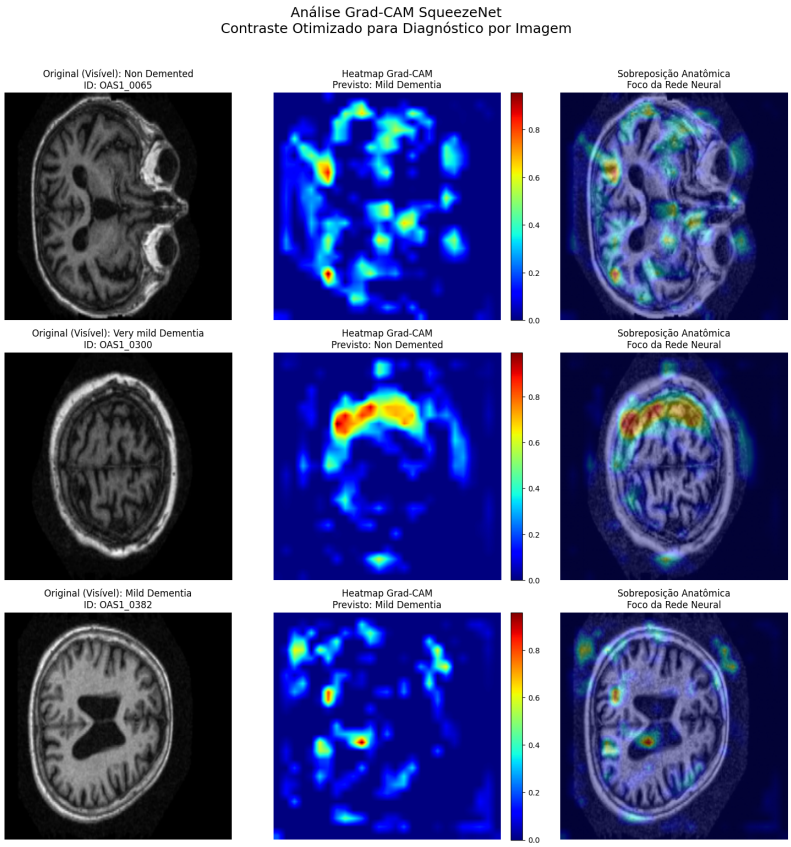


Figura 5.6 – Mapa Grad-CAM do modelo *SqueezeNet*

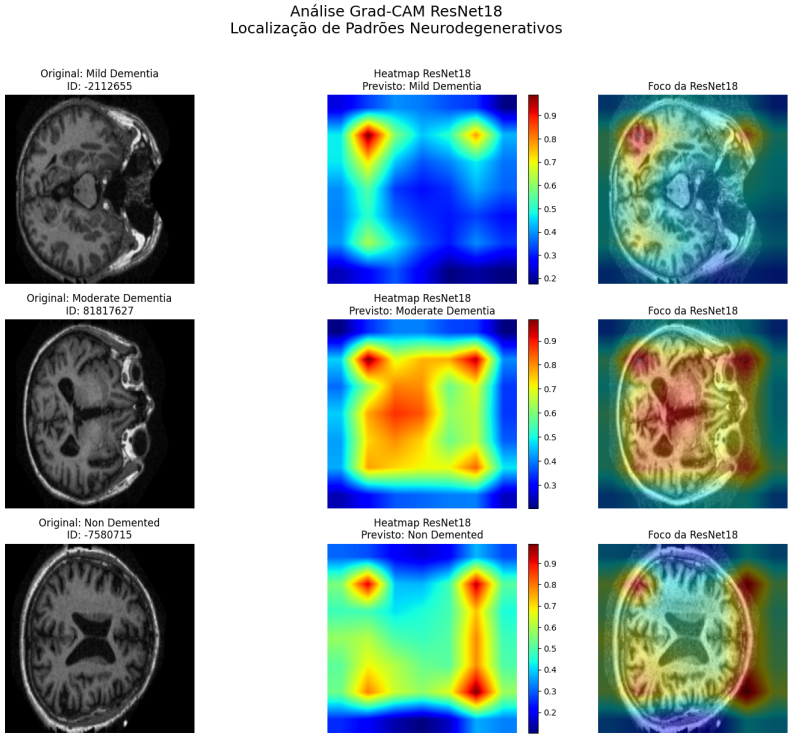


Figura 5.7 – Mapa Grad-CAM do modelo *ResNet18*

Os elevados índices de acurácia observados, especialmente na *ResNet18*, devem ser analisados com cautela devido ao método de balanceamento utilizado. Como o *over-*

sampling da classe *Mild Dementia* foi realizado antes da fragmentação dos dados em treino, validação e teste, ocorreu um cenário de vazamento de dados (*data leakage*): amostras idênticas de um mesmo paciente foram distribuídas entre os diferentes *splits*.

Portanto, a performance superior dessas arquiteturas reflete uma memorização de padrões redundantes, e não necessariamente uma maior capacidade de generalização do modelo em dados inéditos. Esse resultado evidencia que o viés introduzido no pré-processamento influu as métricas finais, mascarando o real desempenho das redes frente ao desbalanceamento severo do dataset original.

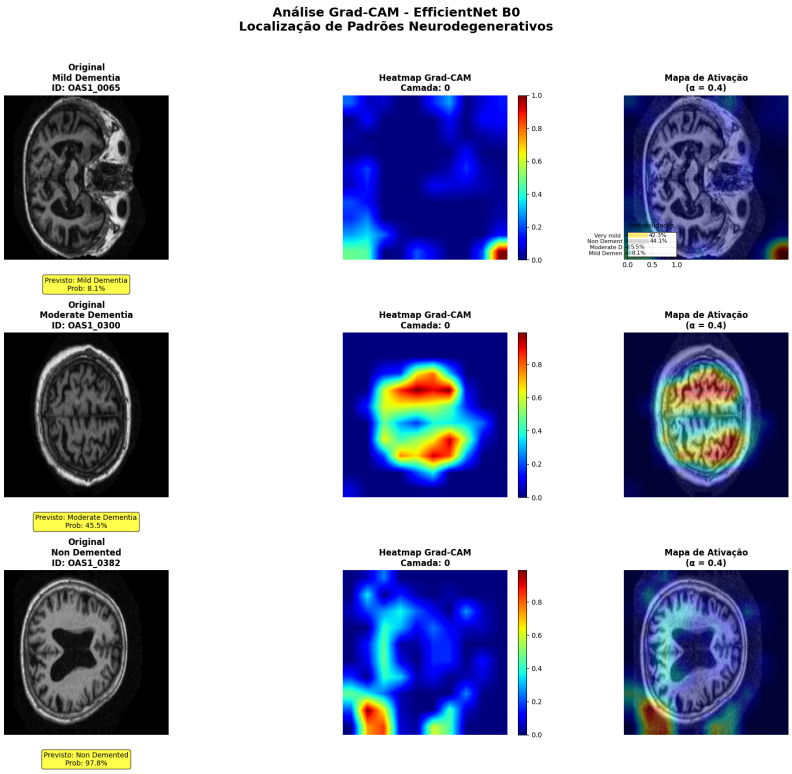


Figura 5.8 – Mapa Grad-CAM do modelo *EfficientNet-B3*

A Figura 5.8 mostra o mapa Grad-CAM do modelo *EfficientNet-B3*, evidenciando as regiões de maior ativação utilizadas na tomada de decisão. Observa-se que, para a classe *Non Demented*, o modelo apresentou alta confiança (97,8%), enquanto nas classes *Mild Dementia* e *Moderate Dementia* houve maior incerteza e sobreposição de probabilidades, refletindo a dificuldade do modelo em distinguir estágios intermediários da doença.

5.2 Conv3D

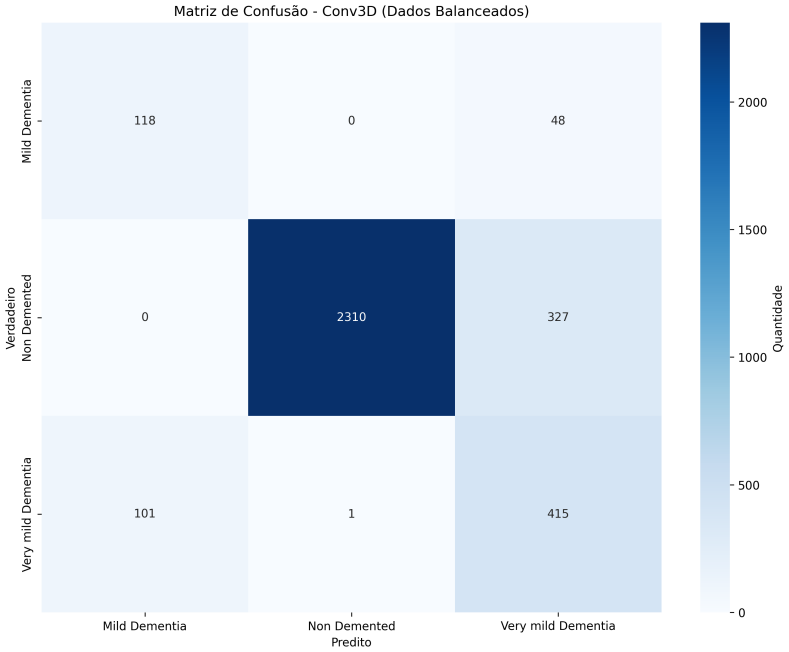


Figura 5.9 – Matix de confusão do modelo *Conv3D*

O modelo Conv3D foi treinado por até 25 épocas (com *early stopping*) utilizando 242 pacientes no treino, 53 na validação e 52 no teste, sem *data leakage*. Apesar do grande número de volumes gerados a partir das imagens, o número de pacientes por classe foi reduzido e fortemente desbalanceado, especialmente para *Moderate Dementia* (apenas 2 pacientes no dataset completo).

No conjunto de teste, o modelo apresentou acurácia geral de 21,63% e acurácia balanceada de 25,36%. O *F1-Score Macro* foi de 0,1498 e o *F1-Score Weighted* de 0,2912, enquanto o coeficiente de correlação de Matthews (MCC) foi de 0,0579, indicando correlação muito fraca entre as predições e os rótulos reais. Esses valores evidenciam baixa capacidade de generalização.

A análise por classe revela comportamento assimétrico. A classe *Mild Dementia* apresentou recall elevado (0,764), porém com baixa precisão (0,11), resultando em alto número de falsos positivos. A classe *Non Demented* obteve alta precisão (0,98), mas recall reduzido (0,215), indicando que o modelo foi excessivamente conservador ao predizer essa categoria. Já *Very Mild Dementia* apresentou desempenho bastante limitado (precisão de 0,07 e recall de 0,036), demonstrando dificuldade na identificação de alterações sutis. A classe *Moderate Dementia* não esteve presente no conjunto de teste, o que impossibilitou sua avaliação nessa etapa.

De forma geral, os resultados indicam que, mesmo com estratégias de *oversampling* no treino e uso de *sampler* balanceado, o forte desbalanceamento entre pacientes e o número reduzido de exemplos nas classes minoritárias limitaram o aprendizado do modelo. Assim, considerando os baixos valores de acurácia, *F1-Score* e MCC, o modelo ainda não apresenta desempenho adequado para aplicação clínica, sendo necessários conjuntos de dados mais robustos e melhor balanceados para avanços futuros.

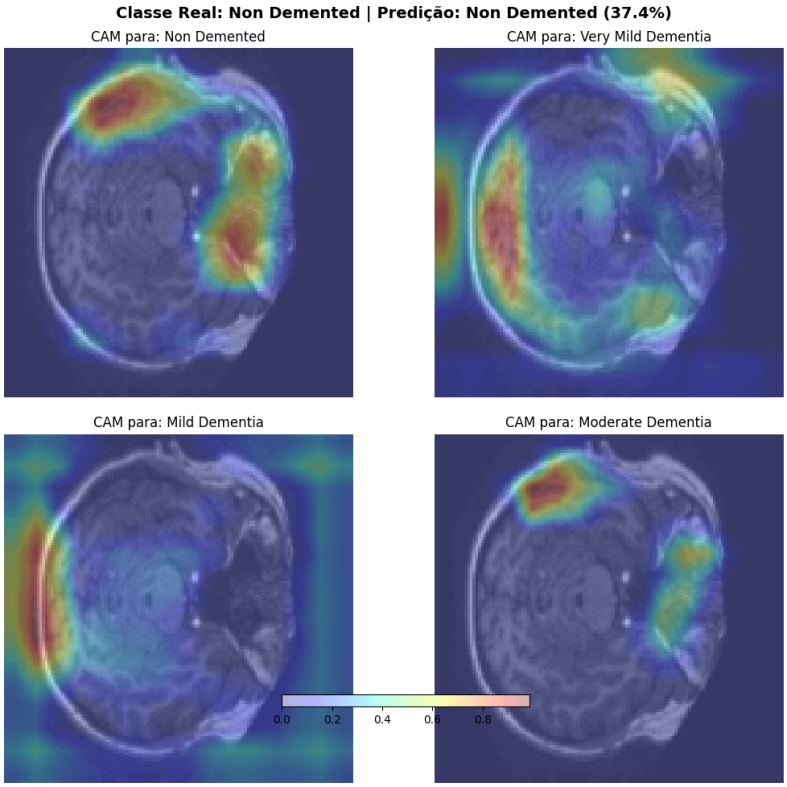


Figura 5.10 – Classe verdadeira: *Non Demented*, Predição: *Non Demented*

As Figuras 5.10, 5.11 e 5.12 apresentam os mapas de ativação Grad-CAM gerados pela rede *Conv3D*. Por se tratar de um modelo tridimensional que processa volumes completos, a análise não é realizada em imagens isoladas, mas sim no volume cerebral completo do paciente. Diferentemente de abordagens 2D que analisam fatias independentemente, o modelo *Conv3D* é capaz de explorar correlações espaciais tridimensionais e padrões volumétricos ao longo de todo o encéfalo, incluindo regiões que podem apresentar ou não biomarcadores da doença de Alzheimer. Dessa forma, a ativação destacada nos mapas Grad-CAM reflete as regiões cerebrais que foram mais determinantes para a decisão do modelo, considerando a estrutura tridimensional completa do cérebro.

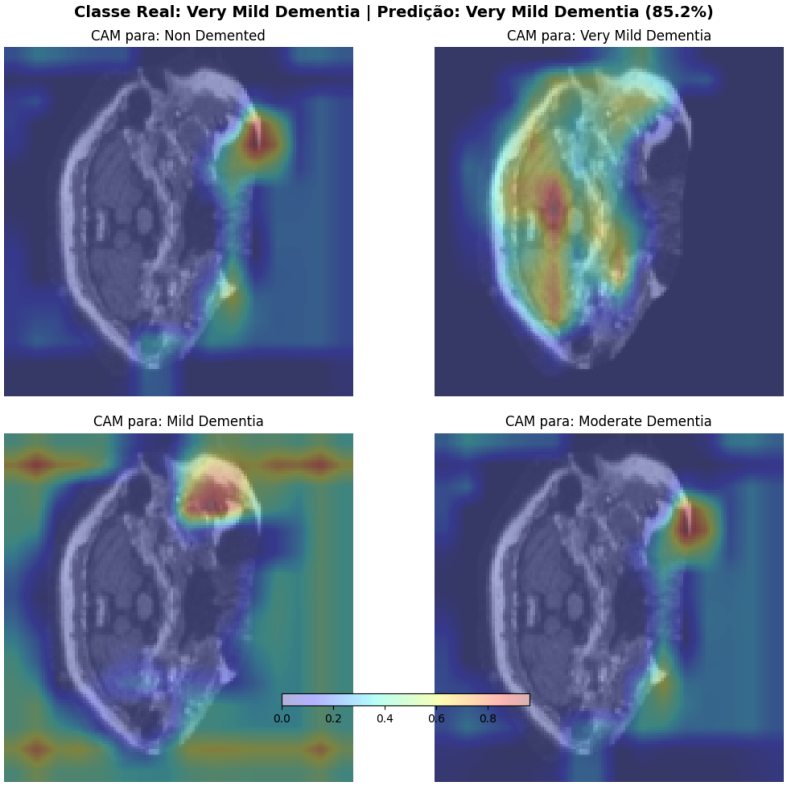


Figura 5.11 – Classe verdadeira: *Very mild Dementia*, Predição: *Very mild Dementia*

Conforme ilustrado na Figura 5.11, as imagens periféricas do conjunto de dados foram incluídas no treinamento. O descarte dessas camadas permitiria que o modelo priorizasse a análise das áreas cerebrais sabidamente afetadas pelo Alzheimer, otimizando o aprendizado (BRAAK; BRAAK, 1991).

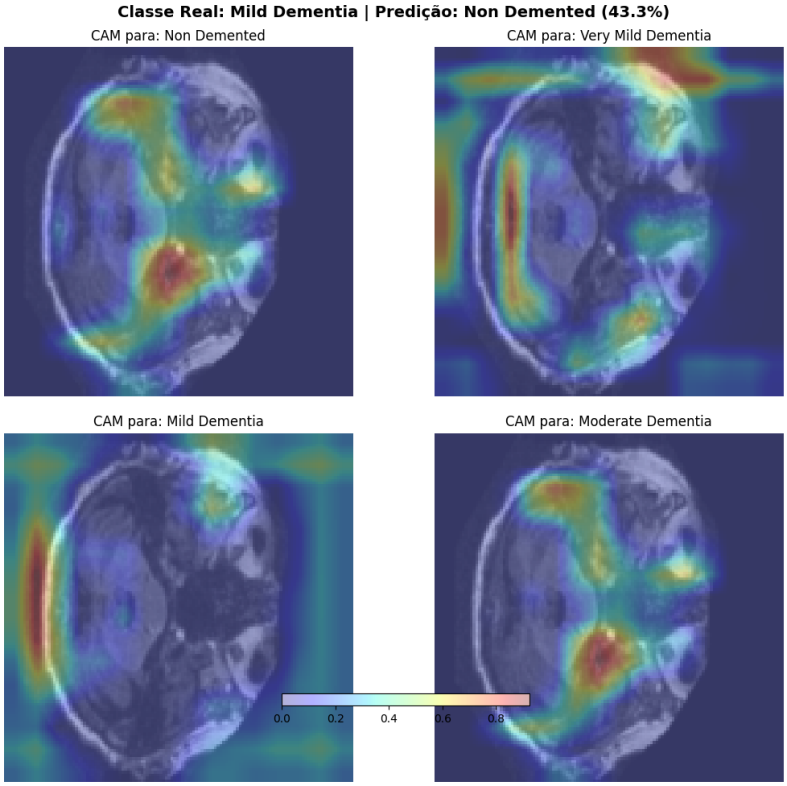


Figura 5.12 – Classe verdadeira: *Non Demented*, Predição: *Very mild Dementia*

Apesar de o modelo *Conv3D* ter alcançado uma acurácia geral de 0,80, a análise por classe evidencia limitações relevantes. O desempenho para o diagnóstico de *Mild Dementia* foi particularmente insatisfatório, com recall de apenas 0,09, o que indica que o modelo praticamente não reconhece esses casos. Esse resultado está diretamente relacionado ao forte desbalanceamento do conjunto de dados, no qual a classe *Non Demented* representa cerca de 80% das amostras. Consequentemente, o modelo tende a favorecer previsões para as classes majoritárias, elevando a acurácia global, mas reduzindo sua capacidade de identificar estágios iniciais e intermediários da doença. Assim, a acurácia isolada se mostra inadequada como métrica principal, reforçando a importância de considerar métricas por classe e de aplicar técnicas de balanceamento para aprimorar o poder discriminativo do modelo.

A arquitetura *AlexNet* não apresentou robustez satisfatória diante do desequilíbrio severo dos dados médicos. Verificou-se que a ausência de amostras da classe *Moderate Dementia* em determinados lotes (*mini-batches*) resultava em um erro nulo para essa categoria, culminando na estagnação ou no colapso do gradiente. Para mitigar esse fenômeno e impedir que a rede neural negligenciasse as classes minoritárias durante o processo de otimização, foi imperativo o emprego de técnicas de *oversampling* (BUDA; MAKI; MAZUROWSKI, 2018).

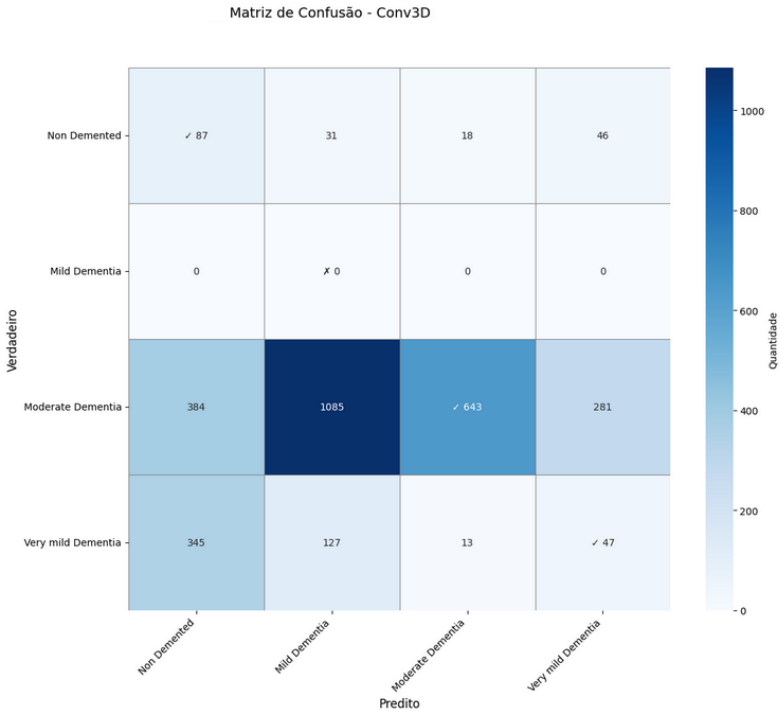


Figura 5.13 – Matriz de confusão do Modelo *Conv3d*

A Figura 5.13 mostra a Matriz de confusão do Modelo *Conv3D*, foi obtido esse resultado com *Oversampling* apenas no treino e no calulo dos pesos tambem consegui uma estratificação por Paciente sem *Data Leakage*. Estudos ((KAPOOR; NARAYANAN, 2023),(SEH et al., 2020),)sugerem que o vazamento não intencional de data leakage em trabalhos médicos.

5.3 Resumo

Os modelos AlexNet e SqueezeNet não apresentaram bom desempenho, demonstrando não serem robustos o suficiente para dados médicos, além de apresentarem risco de vazamento de dados entre os experimentos. O ResNet18 obteve bom desempenho, mostrando-se adequado para a análise de imagens médicas.

Por se tratar de imagens de ressonância magnética, que são inerentemente um problema 3D, e considerando que os slices (fatias) possuem uma ordem sequencial, foi utilizado o modelo Conv3D. No entanto, esse modelo não contava com pré-treinamento em grandes bases de dados, o que justifica seu baixo desempenho quando comparado a abordagens com transfer learning.

6 Conclusão

Com o aumento da quantidade de dados provenientes de exames de ressonância magnética, observa-se um crescimento significativo nos custos associados à análise desses exames. Além disso, o processo de interpretação é altamente dependente do trabalho dos radiologistas, sendo uma atividade demorada e cognitivamente cansativa.

Os modelos de redes neurais convolucionais mostraram-se adequados para a tarefa de detecção da Doença de Alzheimer, considerando o tamanho limitado da base de dados disponível. Observou-se que o uso de modelos pré-treinados elevou significativamente o desempenho das abordagens testadas, mesmo com um número reduzido de épocas de treinamento. Entretanto, a análise de explicabilidade revelou que, em alguns casos, o modelo concentrou sua atenção em regiões fora das áreas anatômicas de interesse, indicando a necessidade de aprimoramentos nos métodos de pré-processamento, segmentação ou seleção de características. Esses resultados sugerem que, embora a abordagem seja promissora, ainda há espaço para avanços no que diz respeito à robustez e à interpretabilidade dos modelos.

No caso de conjuntos de dados com muitas imagens semelhantes, como o OASIS-1, arquiteturas mais leves, como a *ResNet18*, demonstraram desempenho consistente. Em contraste, redes mais profundas, como a *ResNet50* e a *EfficientNet-B3*, apresentam um número significativamente maior de parâmetros, o que aumenta o risco de *overfitting* quando aplicadas a bases de pequeno volume. Dessa forma, a escolha da arquitetura deve considerar cuidadosamente o equilíbrio entre a capacidade representacional do modelo e o tamanho do conjunto de dados, de modo a evitar o sobreajuste e maximizar o desempenho geral.

Considerando a dimensão limitada do conjunto de dados disponível, optou-se por não utilizar arquiteturas baseadas integralmente em *Transformers*, as quais demandam volumes massivos de dados para uma convergência robusta. Para trabalhos futuros, com a expansão da base de dados ou o uso de técnicas avançadas de *pre-training*, torna-se

viável a aplicação de modelos como *Vision Transformer* (ViT) ou *Swin Transformer*. Tais arquiteturas permitiriam explorar correlações espaciais globais e, no caso da *Swin Transformer*, a análise volumétrica 3D, integrando a informação de múltiplos cortes anatômicos de forma simultânea.

6 Referências

AGGARWAL, R. et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. **NPJ digital medicine**, Nature Publishing Group UK London, v. 4, n. 1, p. 65, 2021.

ALABI, M. Ai-assisted medical diagnosis using deep learning and computer vision. 2025.

AWANG, M. K.; ALI, G.; FAHEEM, M. Deep learning techniques for alzheimer's disease detection in 3d imaging: A systematic review. **Health Science Reports**, Wiley Online Library, v. 7, n. 9, p. e70025, 2024.

BARGAGNA, F. et al. Bayesian convolutional neural networks in medical imaging classification: a promising solution for deep learning limits in data scarcity scenarios. **Journal of Digital Imaging**, Springer, v. 36, n. 6, p. 2567–2577, 2023.

BRAAK, H.; BRAAK, E. Neuropathological stageing of alzheimer-related changes. **Acta neuropathologica**, Springer, v. 82, n. 4, p. 239–259, 1991.

BRUNO, J. B. d. S. L.; SIQUEIRA, R. N. S. et al. Aa utilização de ressonância magnética no auxílio do diagnóstico da doença de alzheimer. **Revista Coopex.**, v. 13, n. 1, p. 1–12, 2022.

BUDA, M.; MAKI, A.; MAZUROWSKI, M. A. A systematic study of the class imbalance problem in convolutional neural networks. **Neural networks**, Elsevier, v. 106, p. 249–259, 2018.

CARREIRA, J.; ZISSERMAN, A. Quo vadis, action recognition? a new model and the kinetics dataset. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2017. p. 6299–6308.

CASTRO, M. T. d. et al. Redes neurais convolucionais aplicadas a identificação e classificação de galáxias. 2025.

COELHO, P. M. Periocular efficientnet: um modelo profundo para reconhecimento periocular. 2024.

FEITOSA, J. d. C. Análise da explicabilidade dos métodos de inteligência artificial explicável aplicados a segmentação de imagens de peixes pacu. Universidade Estadual Paulista (Unesp), 2024.

FILHO, R. M. R.; VIDAL, M. C. Classificação de imagens médicas de alzheimer utilizando redes neurais convolucionais para a identificação da doença. 2024.

GEGENFURTNER, A. et al. The challenges of studying visual expertise in medical image diagnosis. **Medical Education**, Wiley Online Library, v. 51, n. 1, p. 97–104, 2017.

GOMIDES, A. V. M. Análise do impacto do aumento de dados no reconhecimento de ferrugem em folhas de café. Universidade Federal de Uberlândia, 2024.

HASSAN, N. et al. Stacked cnn-based multichannel attention networks for alzheimer disease detection. **Scientific Reports**, Nature Publishing Group UK London, v. 15, n. 1, p. 5815, 2025.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. [S.l.]: Springer Science & Business Media, 2009.

HE, K. et al. Deep residual learning for image recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 770–778.

HOWARD, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv:1704.04861**, 2017.

HUANG, G. et al. Densely connected convolutional networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 4700–4708.

IANDOLA, F. N. et al. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. **arXiv preprint arXiv:1602.07360**, 2016.

IANDOLA, F. N. et al. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. **CoRR**, abs/1602.07360, 2016. Disponível em: <http://arxiv.org/abs/1602.07360>.

JI, S. et al. 3d convolutional neural networks for human action recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 35, n. 1, p. 221–231, 2012.

JR, C. R. J. et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. **Journal of Magnetic Resonance Imaging**, v. 27, n. 4, p. 685–691, 2008.

JUNIOR, A. M. d. Q. Classificação automática da doença de alzheimer por meio de processamento de imagens e algoritmos de inteligência artificial. Universidade Federal do Rio Grande do Norte, 2023.

KAPOOR, S.; NARAYANAN, A. Leakage and the reproducibility crisis in machine-learning-based science. **Patterns**, Elsevier, v. 4, n. 9, 2023.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in Neural Information Processing Systems (NeurIPS)**. [S.l.: s.n.], 2012. v. 25, p. 1097–1105.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, Ieee, v. 86, n. 11, p. 2278–2324, 2002.

LI, M. et al. Medical image analysis using deep learning algorithms. **Frontiers in public health**, Frontiers Media SA, v. 11, p. 1273253, 2023.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. **Advances in neural information processing systems**, v. 30, 2017.

MARCHIORI, C. E. A. et al. Aplicação de uma rede neural de convolução para posicionamento de seccionadora em linha de produção. Blumenau, SC., 2023.

MARCUS, D. S. et al. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented adults. **Journal of cognitive neuroscience**, v. 19, n. 9, p. 1498–1507, 2007.

MCDONALD, R. J. et al. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. **Academic radiology**, Elsevier, v. 22, n. 9, p. 1191–1198, 2015.

MÜLLER, D.; SOTO-REY, I.; KRAMER, F. An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks. **Ieee Access**, IEEE, v. 10, p. 66467–66480, 2022.

PADOVESE, B. T. Suporte ao diagnóstico da doença de alzheimer a partir de imagens de ressonância magnética. Universidade Estadual Paulista (Unesp), 2017.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. "why should i trust you?"explaining the predictions of any classifier. In: **Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining**. [S.l.: s.n.], 2016. p. 1135–1144.

RODRIGUES, M. E. S. Avaliação de desempenho de redes neurais convolucionais no diagnóstico de amiloidose cardíaca. Universidade Federal da Paraíba, 2024.

SANTOS, S. S. et al. Interpretable diagnosis of skin cancer using deep learning. In: **Congresso Brasileiro de Automática-CBA**. [S.l.: s.n.], 2022. v. 3, n. 1.

SEH, A. H. et al. Healthcare data breaches: insights and implications. In: MDPI. **Healthcare**. [S.l.], 2020. v. 8, n. 2, p. 133.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of big data**, Springer, v. 6, n. 1, p. 1–48, 2019.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.

SOUSA, C. E. P. et al. Eficiência de redes neurais convolucionais na identificação de edifícios em condições desafiadoras de imagens. **OBSERVATÓRIO DE LA ECONOMÍA LATINOAMERICANA**, v. 23, n. 5, p. e10080–e10080, 2025.

SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 2818–2826.

TAKAHASHI, S. et al. Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review. **Journal of Medical Systems**, Springer, v. 48, n. 1, p. 84, 2024.

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.

TORCHVISION.MODELS.EFFICIENTNET_B3 — Torchvision documentation. Disponível em: <https://docs.pytorch.org/vision/main/models/generated/torchvision.models.efficientnet_b3.html>.

TRAN, D. et al. Learning spatiotemporal features with 3d convolutional networks. In: **Proceedings of the IEEE International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2015. p. 4489–4497.

TURANLI, M. Multi-orientation hippocampus-centered 3d cnn with attention mechanism for alzheimer's disease classification from mri scans. **medRxiv**, Cold Spring Harbor Laboratory Press, p. 2025–06, 2025.

VAROQUAUX, G.; COLLIOT, O. Evaluating machine learning models and their diagnostic value. **Machine Learning for Brain Disorders**, Springer, p. 601–630, 2023.

VO, T. N.; HO, B. N. 3d brain mri classification for alzheimer diagnosis using cnn with data augmentation. **arXiv preprint arXiv:2505.04097**, 2025.

WAITE, S. et al. Interpretive error in radiology. **American Journal of Roentgenology**, American Roentgen Ray Society, v. 208, n. 4, p. 739–749, 2017.

WEN, J. et al. Convolutional neural networks for alzheimer's disease classification: From pride to prejudice. **Medical Image Analysis**, Elsevier, v. 63, p. 101694, 2020.

WEN, J. et al. Convolutional neural networks for classification of alzheimer's disease: overview and reproducible evaluation. **Medical image analysis**, Elsevier, v. 63, p. 101694, 2020.

ZHANG, W. et al. An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms. **Medical Physics**, Wiley Online Library, v. 23, n. 4, p. 595–601, 1996.