

Mateus Rosolem Baroni

**Diagnóstico Automatizado da Doença de
Alzheimer Utilizando Redes Neurais
Convolucionais**

São José do Rio Preto

2025

Mateus Rosolem Baroni

Diagnóstico Automatizado da Doença de Alzheimer Utilizando Redes Neurais Convolucionais

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho” como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Universidade Estadual Paulista “Júlio de Mesquita Filho”
Instituto de Biociências, Letras e Ciências Exatas
Departamento de Ciência de Computação e Estatística

Orientador: Prof. Dr. Wallace Correa de Oliveira Casaca

São José do Rio Preto
2025

B266d

Baroni, Mateus Rosolem

Diagnóstico automatizado da doença de Alzheimer utilizando redes neurais convolucionais / Mateus Rosolem Baroni. -- São José do Rio Preto, 2025
51 p. : tabs., fotos

Trabalho de conclusão de curso (Bacharelado - Ciência da Computação) -
Universidade Estadual Paulista (UNESP), Instituto de Biociências Letras e Ciências
Exatas, São José do Rio Preto

Orientador: Prof. Dr. Wallace Correa de Oliveira Casaca

1. Doença de Alzheimer. 2. Redes Neurais Convolucionais. 3. Diagnóstico
Automatizado. 4. Imagens Médicas. 5. Deep Learning. I. Título.

“A melhor forma de prever o futuro é criá-lo.” Dedico este trabalho a quem acreditou no meu futuro, ou seja, a Deus, meu João Paulo Baroni e minha mãe Rosany Rosolem, toda minha família e pessoas queridas que conheci na jornada até aqui, mesmo quando ele ainda era só uma ideia.

Agradecimentos

Espaço para agradecer formalmente às pessoas e instituições que contribuíram para o trabalho. Ao orientador, Prof. Dr. Wallace Correa de Oliveira Casaca, pela orientação e apoio. Aos membros da banca examinadora, pela disponibilidade e valiosas contribuições. À família e amigos, pelo suporte ao longo da jornada acadêmica. À UNESP, pelo ambiente de aprendizado e infraestrutura. Também aos meus amigos Leandro, Luan, Otávio, Yhan e por toda companhia e apoio.

“Não são as respostas que movem o mundo, são as perguntas.” Agradeço aos professores que me ensinaram a questionar antes de aceitar.

Resumo

A Doença de Alzheimer (DA) representa um desafio crescente para a saúde pública global, demandando métodos de diagnóstico mais rápidos e precisos. Este trabalho propõe um sistema automatizado para o diagnóstico da DA utilizando diferentes arquiteturas de Deep Learning aplicadas a imagens de Ressonância Magnética (MRI). O estudo visa classificar pacientes entre Controle Normal (NC), Comprometimento Cognitivo Leve (CCL) e DA, através da comparação sistemática de seis arquiteturas profundas de visão, incluindo uma Rede Neural Convolucional (CNN) customizada, modelos residuais (ResNet), modelos eficientes (EfficientNet), Vision Transformers (ViT) e Redes de Cápsulas (CapsNet). O desempenho é avaliado por um conjunto robusto de métricas como acurácia, sensibilidade, especificidade e F1-score. Resultados indicam desempenho competitivo e o forte potencial destas arquiteturas como ferramentas de apoio ao diagnóstico clínico, oferecendo um comparativo robusto para pesquisas futuras.

Palavras-chave: Doença de Alzheimer; Redes Neurais Convolucionais; Diagnóstico Automático; Imagens Médicas; Deep Learning.

Abstract

Alzheimer's Disease (AD) is a growing global public health challenge requiring faster and more accurate diagnostics. This work proposes an automated system for AD diagnosis using various Deep Learning architectures applied to Magnetic Resonance Imaging (MRI) scans. The study aims to classify patients into multiple diagnostic categories, including Normal Control (NC), Mild Cognitive Impairment (MCI), and AD, through the systematic comparison of six state-of-the-art architectures, including a customized Convolutional Neural Network (CNN), residual networks (ResNet), efficient models (EfficientNet), Vision Transformers (ViT), and Capsule Networks (CapsNet). Performance is evaluated using a robust set of metrics such as accuracy, sensitivity, specificity, and F1-score. Results show competitive performance and the strong potential of these architectures as clinical decision support tools, providing a robust benchmark for future research.

Keywords: Alzheimer's Disease; Convolutional Neural Networks; Automatic Diagnosis; Medical Imaging; Deep Learning.

Lista de ilustrações

Figura 1 – Exames de ressonância magnética (MRI) de (a) um cérebro normal e (b) um cérebro com doença de Alzheimer. (GOENKA; TIWARI, 2021)	13
Figura 2 – Arquitetura de uma Rede Neural Convolutacional (CNN) clássica	20
Figura 3 – Arquitetura da Rede Neural Convolutacional ResNet	20
Figura 4 – Arquitetura da EfficientNet-B0 com blocos MBConv	21
Figura 5 – Exemplos de Fatias Axiais do Dataset MIRIAD	30
Figura 6 – Pipeline Completo de Pré-processamento e Treinamento dos Dados . .	31
Figura 7 – Desempenho das Estratégias de <i>fine-tuning</i> no Dataset ADNI (Etapa 2 - Classificação Multiclasse de 4 Categorias). Linhas tracejadas indicam o desempenho da Etapa 1. A estratégia Full Fine-Tune apresentou desempenho excepcional: 99.0% para EfficientNet-V2-M e 92.1% para ResNet152, demonstrando a eficácia do ajuste completo das camadas mesmo em tarefa multiclasse desbalanceada.	40
Figura 8 – Desempenho das Estratégias de <i>fine-tuning</i> no Dataset MIRIAD (Etapa 2 - Classificação Binária AD vs HC). Linhas tracejadas indicam o desempenho da Etapa 1. A estratégia Top Layers com ResNet152 alcançou o melhor resultado geral (83.2%), superando marginalmente a Etapa 1.	41

Lista de tabelas

Tabela 1 – Distribuição de Amostras (Fatias 2D) do Dataset ADNI por Classe . . .	28
Tabela 2 – Hiperparâmetros e Configurações de Otimização.	34
Tabela 3 – Bibliotecas Python Auxiliares	35
Tabela 4 – Resultados da Etapa 0 (Linha de Base) no Dataset MIRIAD	37
Tabela 5 – Resultados da Etapa 0 (Linha de Base) no Dataset ADNI	37
Tabela 6 – Resultados da Etapa 1 (Otimização Avançada) no Dataset MIRIAD .	38
Tabela 7 – Resultados da Etapa 1 (Otimização Avançada) no Dataset ADNI . . .	39
Tabela 8 – Resultados da Etapa 2 (Estratégias de <i>fine-tuning</i>) no Dataset ADNI	39
Tabela 9 – Resultados da Etapa 2 (Estratégias de <i>fine-tuning</i>) no Dataset MIRIAD	40
Tabela 10 – Comparação de Desempenho com Trabalhos Relacionados	44

Lista de Abreviaturas e Siglas

ADNI Alzheimer's Disease Neuroimaging Initiative

AUC-ROC Área sob a Curva ROC

BN Batch Normalization

CapsNet Capsule Network (Rede de Cápsulas)

CNN Convolutional Neural Network (Rede Neural Convolucional)

DA Doença de Alzheimer

FN Falsos Negativos

FP Falsos Positivos

HC Controle Saudável (Healthy Control)

IA Inteligência Artificial

MCI/CCL Mild Cognitive Impairment / Comprometimento Cognitivo Leve

MIRIAD Minimal Interval Resonance Imaging in Alzheimer's Disease

MRI Magnetic Resonance Imaging (Imagem por Ressonância Magnética)

NC Controle Normal (Normal Control)

OASIS Open Access Series of Imaging Studies

PET Tomografia por Emissão de Pósitrons

ReLU Rectified Linear Unit

ResNet Residual Network

RNA Redes Neurais Artificiais

VGG Visual Geometry Group

ViT Vision Transformer

VN Verdadeiros Negativos

VP Verdadeiros Positivos

Sumário

	Lista de ilustrações	7
	Lista de tabelas	8
	Sumário	10
1	INTRODUÇÃO	12
1.1	Contextualização	12
1.2	Justificativa	13
1.3	Problema de Pesquisa	13
1.4	Hipótese	14
1.5	Objetivos	15
1.6	Contribuições Alcançadas	15
1.7	Organização do Trabalho	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	A Doença de Alzheimer	17
2.2	Métodos Diagnósticos Tradicionais	17
2.3	Neuroimagem na DA	18
2.3.1	Princípios da Imagem por Ressonância Magnética (MRI)	18
2.3.2	Inteligência Artificial Aplicada ao Alzheimer	18
2.3.3	Vantagens, Desafios e Limitações	18
2.4	Redes Neurais Convolucionais e Arquiteturas	19
2.4.1	Conceitos Fundamentais	19
2.4.2	Arquiteturas Tradicionais	19
2.4.3	Aprendizagem por Transferência	21
3	TRABALHOS RELACIONADOS	23
4	METODOLOGIA	25
4.1	Bases de Dados	25
4.1.1	MIRIAD	25
4.1.2	ADNI	26
4.2	Pré-processamento	29
4.2.1	Normalização e Padronização de Intensidade	29

4.2.2	Redimensionamento	30
4.2.3	Aumento de Dados (Data Augmentation)	31
4.2.4	Divisão dos Dados e Validação Cruzada	32
4.3	Arquiteturas dos Modelos e Estratégia de Treinamento	32
4.3.1	Detalhamento das Arquiteturas Implementadas	32
4.3.2	Configuração das Camadas de Saída	33
4.3.3	Hiperparâmetros e Otimização	34
4.4	Ambiente	34
4.5	Métricas de Avaliação	34
5	RESULTADOS	36
5.1	Análise Exploratória dos Dados	36
5.2	Análise de Desempenho por Estágio	36
5.2.1	Etapa 0: Linha de Base	36
5.2.2	Etapa 1: Otimização de Parâmetros e Técnicas de Ajuste	38
5.2.3	Etapa 2: Estratégias de <i>fine-tuning</i>	39
5.2.3.1	Análise Comparativa dos Resultados de Fine-Tuning	39
5.2.3.2	Divergência de Comportamento entre Datasets	42
5.2.3.3	Implicações Práticas e Recomendações	42
5.3	Comparação com a Literatura	44
6	CONCLUSÃO	46
6.1	Síntese dos Resultados e Resgate dos Objetivos	46
6.2	Propostas para Trabalhos Futuros	47
	REFERÊNCIAS	48

1 Introdução

O avanço das tecnologias de computação e a crescente disponibilidade de dados médicos têm impulsionado uma transformação significativa na área da saúde, especialmente no campo do diagnóstico por imagem. A integração entre a Inteligência Artificial e a prática clínica oferece oportunidades promissoras para aumentar a precisão e a eficiência na detecção de patologias complexas. Este trabalho se insere nesse contexto de inovação, investigando como métodos computacionais modernos podem auxiliar no enfrentamento de desordens neurodegenerativas que impactam milhões de pessoas.

1.1 Contextualização

A Doença de Alzheimer (DA) representa um dos maiores desafios de saúde pública do século XXI (Alzheimer's Association, 2024; LANE; HARDY; SCHOTT, 2018; MASTERS et al., 2015). Caracterizada como uma desordem neurodegenerativa progressiva e irreversível, a DA afeta predominantemente a população idosa, comprometendo gradualmente funções cognitivas essenciais como a memória, o raciocínio e a linguagem.

Com o envelhecimento populacional em escala global, a predominância da doença tem aumentado exponencialmente, gerando um impacto social e econômico devastador para pacientes, famílias e sistemas de saúde. A complexidade de sua patologia, que envolve o acúmulo de placas beta-amiloide e emaranhados neurofibrilares, ainda é objeto de intensa pesquisa, e a ausência de uma cura torna o diagnóstico precoce e preciso uma ferramenta indispensável para o manejo clínico e a busca por terapias que possam retardar sua progressão. (TUFAIL; MA; ZHANG, 2018) (FENG et al., 2019) (GOEDERT et al., 1991) (JACK et al., 2018)

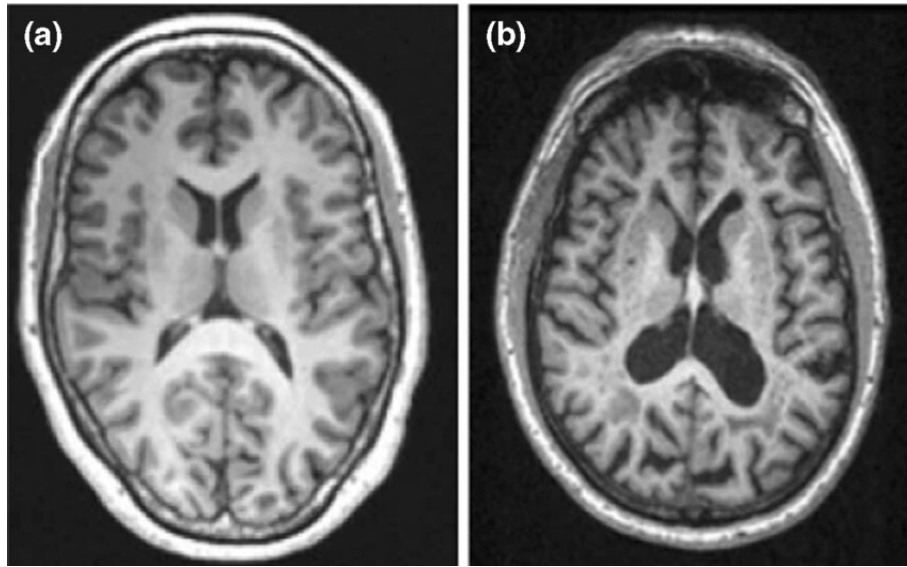


Figura 1 – Exames de ressonância magnética (MRI) de (a) um cérebro normal e (b) um cérebro com doença de Alzheimer. (GOENKA; TIWARI, 2021)

1.2 Justificativa

O diagnóstico precoce da Doença de Alzheimer é um fator crítico que influencia diretamente na qualidade de vida do paciente. Uma identificação da doença em seus estágios iniciais, como durante a fase de Comprometimento Cognitivo Leve (CCL), permite a implementação de intervenções farmacológicas e não farmacológicas que podem mitigar os sintomas e retardar o declínio cognitivo.

Além disso, um diagnóstico assertivo possibilita que pacientes e familiares se preparem para os desafios futuros, tomando decisões importantes sobre cuidados e planejamento de vida. Nesse cenário, a interseção entre a neuroimagem e a Inteligência Artificial (IA) surge como uma fronteira promissora. Imagens de Ressonância Magnética (MRI) estrutural, por serem não invasivas e amplamente utilizadas, contêm biomarcadores sutis de atrofia cerebral que são indicativos da DA. (PETRELLA et al., 2024; MISTUR et al., 2009; JACK et al., 2018).

A capacidade de algoritmos de *Deep Learning*, em especial as Redes Neurais Convolucionais (CNNs), de aprender e identificar padrões complexos em dados visuais torna-os ferramentas ideais para analisar essas imagens e auxiliar na detecção de alterações muitas vezes imperceptíveis ao olho humano.

1.3 Problema de Pesquisa

Apesar dos avanços, o diagnóstico definitivo da Doença de Alzheimer ainda depende fortemente de análises *post-mortem* (LANE; HARDY; SCHOTT, 2018; MASTERS

et al., 2015). Os métodos clínicos atuais baseiam-se em uma combinação de avaliações neuropsicológicas, análise do histórico do paciente e exames de imagem, como a MRI e a Tomografia por Emissão de Pósitrons (PET) (JUGANAVAR; JOSHI, 2023; CHAIRI et al., 2024). Contudo, essa abordagem apresenta limitações significativas: os testes cognitivos podem sofrer a influência de fatores como escolaridade e estado emocional (HEMMY et al., 2020); a análise visual de exames de imagem é um processo subjetivo, dependente da experiência do radiologista e demorado (SOROUR et al., 2024; EL-GENEEDY et al., 2023); e marcadores bioquímicos, obtidos por punção lombar, são invasivos e de alto custo (JUGANAVAR; JOSHI, 2023; Alzheimer’s Association, 2024).

A dificuldade em diferenciar o envelhecimento normal do declínio cognitivo patológico em seus estágios iniciais representa o principal gargalo (HEMMY et al., 2020; MISTUR et al., 2009). Diante desse contexto, emerge a seguinte questão de pesquisa:

Qual arquitetura de Deep Learning oferece o melhor equilíbrio entre precisão diagnóstica e generalização para a classificação automatizada da Doença de Alzheimer em ressonâncias magnéticas, particularmente na crítica distinção entre envelhecimento saudável e estágios iniciais de demência?

1.4 Hipótese

A hipótese central deste trabalho é que a aplicação de modelos de *Deep Learning*, notadamente arquiteturas avançadas de visão como Redes Neurais Convolucionais (CNNs) do tipo *Feedforward*, Vision Transformers (ViT), CapsNet, ResNet e EfficientNet, pode superar as limitações dos métodos diagnósticos tradicionais ao extrair e classificar automaticamente biomarcadores de neurodegeneração presentes em imagens de ressonância magnética.

Postula-se que esses modelos, uma vez treinados em um conjunto de dados diversificado como o MIRIAD (MALONE et al., 2013), são capazes de alcançar um alto grau de acurácia, sensibilidade e especificidade na distinção entre cérebros saudáveis e com Alzheimer. Adicionalmente, acredita-se que uma comparação sistemática entre diferentes arquiteturas permitirá identificar quais modelos são mais adequados para essa tarefa específica, e que técnicas de *transfer learning*, como o fine-tuning, podem potencializar ainda mais o desempenho dos algoritmos mais promissores, consolidando sua viabilidade como ferramenta de apoio à decisão clínica.

1.5 Objetivos

O objetivo central desta pesquisa é estabelecer um *framework* comparativo para diagnóstico assistido de Alzheimer utilizando redes neurais profundas e conjuntos de dados públicos de ressonância magnética.

Para cumprir com esse propósito, são definidos os seguintes objetivos específicos:

- Revisar a literatura sobre a fisiopatologia da DA, técnicas de neuroimagem e os fundamentos das arquiteturas de *Deep Learning* (CNN, ViT, ResNet, EfficientNet, CapsNet).
- Estruturar e pré-processar os conjuntos de dados MIRIAD e ADNI, aplicando um pipeline consistente de normalização e aumento de dados.
- Implementar, treinar e validar seis arquiteturas de *Deep Learning* para a tarefa de classificação binária (AD vs. Controle Saudável).
- Avaliar e comparar sistematicamente o desempenho dos modelos em três estágios: (i) linha de base, (ii) com otimização de hiperparâmetros, e (iii) com estratégias de ajuste fino (fine-tuning).
- Analisar criticamente os resultados obtidos, comparando-os com trabalhos correlatos da literatura e discutindo as potencialidades e limitações da abordagem proposta.

1.6 Contribuições Alcançadas

Este trabalho oferece contribuições em múltiplas dimensões, as quais podem ser resumidas da seguinte forma:

- Contribuição Científica: Apresenta um estudo comparativo empírico do desempenho de seis arquiteturas de *Deep Learning* (incluindo modelos convolucionais, baseados em atenção e em cápsulas) no dataset MIRIAD. A validação cruzada com os dados do ADNI fortalece a análise de generalização, oferecendo um benchmark claro para futuras pesquisas.
- Contribuição Tecnológica: Desenvolve um modelo otimizado que serve como protótipo para uma ferramenta de triagem ou auxílio diagnóstico.
- Contribuição Metodológica: Disponibiliza um pipeline de treinamento e avaliação reprodutível, propondo uma análise de desempenho em três estágios que permite isolar e quantificar o impacto de cada etapa de otimização no resultado final.

1.7 Organização do Trabalho

Este Trabalho de Conclusão de Curso está estruturado em seis capítulos. O Capítulo 1 apresenta a introdução ao tema.

O Capítulo 2 aborda a fundamentação teórica sobre a Doença de Alzheimer e as tecnologias de Inteligência Artificial.

O Capítulo 3 revisa os trabalhos relacionados à literatura.

O Capítulo 4 descreve a metodologia detalhada, abrangendo a base de dados, o pré-processamento, as arquiteturas, o ambiente de desenvolvimento e as métricas de avaliação.

O Capítulo 5 apresenta os resultados experimentais em suas três etapas de análise e promove uma discussão aprofundada.

Por fim, o Capítulo 6 expõe as conclusões do trabalho, sintetizando as contribuições e sugerindo direções para pesquisas futuras.

2 Fundamentação Teórica

Este capítulo estabelece as bases conceituais que sustentam a presente pesquisa. Inicialmente, aborda-se a Doença de Alzheimer, sua definição e estágios. Em seguida, são descritos os métodos diagnósticos tradicionais, contrastando-os com o potencial da neuroimagem. Por fim, explora-se a aplicação da Inteligência Artificial neste contexto, detalhando os fundamentos e as arquiteturas profundas de destaque para a execução desta pesquisa.

2.1 A Doença de Alzheimer

A Doença de Alzheimer (DA) é a forma mais comum de demência em idosos, representando um desafio crescente para a saúde pública global. Definida como um transtorno neurodegenerativo progressivo, é caracterizada pelo acúmulo de placas beta-amiloide e emaranhados neurofibrilares de proteína tau no cérebro. Esse processo patológico leva à disfunção sináptica, morte neuronal e atrofia cerebral, resultando em um declínio cognitivo e funcional contínuo (LANE; HARDY; SCHOTT, 2018).

A etiologia da DA é multifatorial, com sintomas iniciais frequentemente sutis, como a perda de memória recente. Com a progressão da doença, surgem dificuldades de linguagem, desorientação e comprometimento do raciocínio. A progressão é comumente dividida em estágios (leve, moderado e grave), culminando em uma dependência completa do paciente (MASTERS et al., 2015). Um desafio diagnóstico crucial reside na fase de Comprometimento Cognitivo Leve (MCI), um estado de transição heterogêneo que pode ser estável (sMCI) ou progredir para a DA (pMCI). A correta identificação dessa trajetória é um dos principais focos da pesquisa em diagnóstico automatizado.

2.2 Métodos Diagnósticos Tradicionais

O diagnóstico da DA em vida é um processo de exclusão, baseado em uma combinação de avaliações clínicas, testes neuropsicológicos e exames de imagem. Testes como o Mini-Exame do Estado Mental (MEEM) são utilizados para quantificar o déficit cognitivo, enquanto a análise do líquido cefalorraquidiano (LCR) pode identificar biomarcadores proteicos, embora seja um procedimento invasivo. Essa abordagem combinada, embora estabelecida, é dependente da experiência do clínico, pode ser demorada e possui limitações na detecção das fases mais incipientes da doença (JACK et al., 2018).

2.3 Neuroimagem na DA

A neuroimagem é uma ferramenta indispensável no diagnóstico da DA, permitindo a visualização *in vivo* das alterações estruturais e funcionais do cérebro.

2.3.1 Princípios da Imagem por Ressonância Magnética (MRI)

A Imagem por Ressonância Magnética (MRI) utiliza um forte campo magnético e ondas de rádio para gerar imagens detalhadas dos tecidos moles do corpo, sem o uso de radiação ionizante. Na DA, a MRI estrutural ponderada em T1 é particularmente útil para identificar padrões de atrofia cerebral.

O termo T1 refere-se ao tempo de relaxamento longitudinal, uma constante de tempo que representa o decaimento exponencial necessário para que os prótons de hidrogênio nos tecidos recuperem a magnetização após serem perturbados por um pulso de radiofrequência. Esse tempo varia conforme a composição molecular e a massa do tecido examinado (por exemplo, gordura versus água), permitindo a geração de contraste na imagem. As alterações mais características observadas incluem a redução do volume do hipocampo e o alargamento dos ventrículos cerebrais, biomarcadores que se correlacionam com a gravidade do declínio cognitivo (FAROOQ et al., 2017).

2.3.2 Inteligência Artificial Aplicada ao Alzheimer

A Inteligência Artificial (IA), especialmente o Aprendizado Profundo (*Deep Learning*) (JO; NHO; SAYKIN, 2019; SOROUR et al., 2024; SHAMRAT et al., 2023), emergiu como uma tecnologia transformadora para a análise de neuroimagens na DA. Algoritmos de *Deep Learning* são capazes de aprender e identificar padrões complexos e sutis de atrofia diretamente das imagens de MRI, que podem ser imperceptíveis ao olho humano. Essa capacidade de extração automática de características posiciona a IA como uma ferramenta poderosa para auxiliar radiologistas e neurologistas no diagnóstico precoce e na diferenciação entre os estágios da doença.

2.3.3 Vantagens, Desafios e Limitações

As principais vantagens da IA no diagnóstico da DA incluem o potencial aumento da precisão, a redução da variabilidade entre observadores e a aceleração do processo de análise. No entanto, sua implementação enfrenta desafios significativos, como a necessidade de grandes volumes de dados de alta qualidade, o risco de sobreajuste (*overfitting*) dos modelos e a dificuldade de interpretação dos resultados, fenômeno conhecido como

caixa-preta (*black box*). A interpretabilidade é uma barreira crítica, pois os clínicos precisam compreender a base da decisão algorítmica para confiar no sistema.

2.4 Redes Neurais Convolucionais e Arquiteturas

2.4.1 Conceitos Fundamentais

As Redes Neurais Convolucionais (CNNs) são uma classe de modelos de *Deep Learning* inspirada no córtex visual humano, que se tornou o padrão-ouro para tarefas de análise de imagens. A arquitetura de uma CNN é composta por uma sequência de camadas especializadas.

As camadas convolucionais aplicam filtros (ou *kernels*) à imagem de entrada através de uma operação de deslizamento controlada por um passo (*stride*). Esses filtros realizam operações matemáticas para criar mapas de características, detectando padrões visuais hierárquicos, desde bordas simples e texturas nas primeiras camadas até formas complexas nas camadas profundas. As funções de ativação, como a ReLU (*Rectified Linear Unit*), introduzem não-linearidade, permitindo que a rede aprenda relações complexas nos dados. As camadas de agrupamento (*pooling*), como o *Max Pooling*, reduzem a dimensionalidade espacial dos mapas de características mantendo as informações mais relevantes, tornando a representação mais robusta a pequenas translações e diminuindo a carga computacional (GOODFELLOW; BENGIO; COURVILLE, 2016).

A Figura 2 ilustra esse fluxo de processamento, demonstrando como a imagem de entrada é transformada sucessivamente através das operações de convolução e subamostragem até a classificação final.

2.4.2 Arquiteturas Tradicionais

Ao longo dos anos, diversas arquiteturas de *Deep Learning* foram propostas e consolidadas. A CNN clássica introduzida por LeCun et al. (1998) estabeleceu os fundamentos das redes convolucionais. A VGG (SIMONYAN; ZISSERMAN, 2015) demonstrou a importância da profundidade da rede através de uma arquitetura simples mas eficaz com camadas convolucionais 3×3 empilhadas.

A ResNet (HE et al., 2016a) revolucionou o campo ao introduzir conexões residuais (*skip connections*) que possibilitaram o treinamento de redes muito mais profundas (com centenas de camadas). Essa inovação permitiu superar o problema de degradação do gradiente (*Vanishing Gradient Problem*), um fenômeno onde os gradientes utilizados para atualizar os pesos da rede tornam-se infinitesimalmente pequenos à medida que retropropagam pelas camadas iniciais, impedindo o aprendizado eficaz em redes profundas.

A arquitetura da ResNet é apresentada na Figura 3, onde é possível observar os blocos residuais característicos desta rede que permitem o fluxo direto da informação.

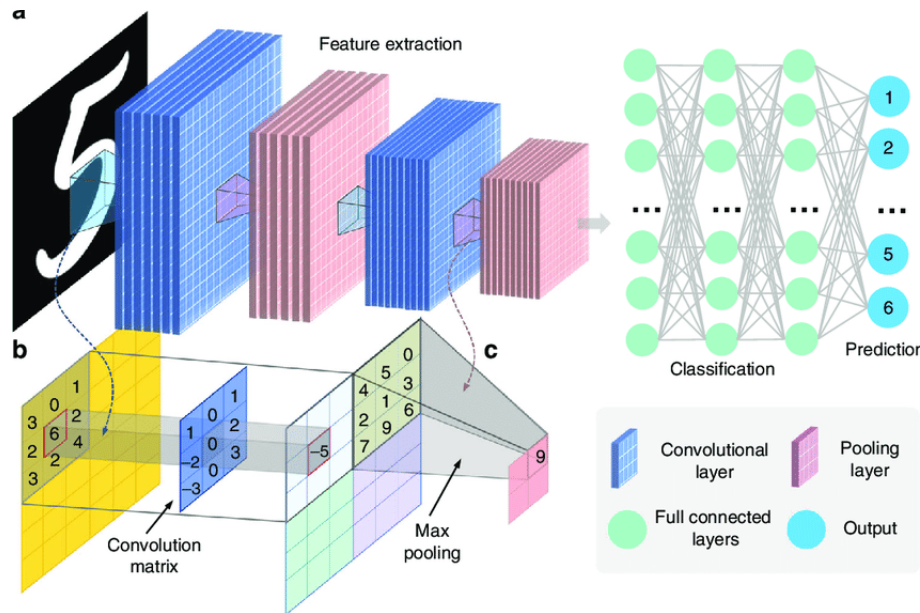


Figura 2 – Arquitetura de uma Rede Neural Convolucional (CNN) clássica

Fonte: Adaptado de Zuo et al. (2022).

A ResNet (HE et al., 2016a) revolucionou o campo ao introduzir conexões residuais (*skip connections*) que possibilitaram o treinamento de redes muito mais profundas, superando o problema de degradação de gradiente. A arquitetura da ResNet é apresentada na Figura 3, onde é possível observar os blocos residuais característicos desta rede.

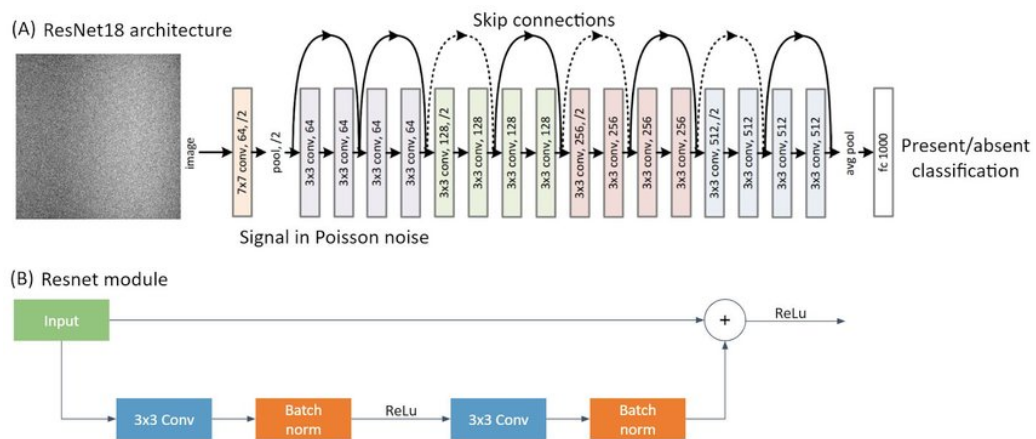


Figura 3 – Arquitetura da Rede Neural Convolucional ResNet

Fonte: Adaptado de (HE et al., 2016b).

A EfficientNet (TAN; LE, 2019a) propôs um método de dimensionamento composto (*compound scaling*) que balanceia profundidade, largura e resolução de forma sistemática, alcançando eficiência superior. A Figura 4 ilustra a arquitetura da EfficientNet-B0

com seus blocos MBConv (Mobile Inverted Bottleneck Convolution) como componentes fundamentais.

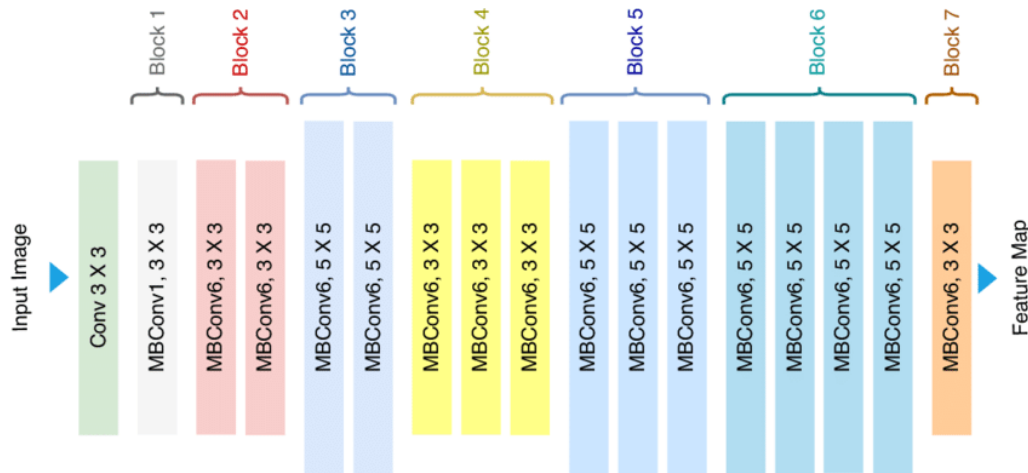


Figura 4 – Arquitetura da EfficientNet-B0 com blocos MBConv

Fonte: Adaptado de (TAN; LE, 2019b).

Para tarefas de segmentação biomédica, a U-Net (RONNEBERGER; FISCHER; BROX, 2015) tornou-se arquitetura de referência devido à sua estrutura encoder-decoder com conexões de salto. Mais recentemente, os Vision Transformers (ViT) (DOSOVITSKIY et al., 2021) adaptaram a arquitetura Transformer para visão computacional, processando imagens como sequências de patches e capturando relações globais através de mecanismos de autoatenção.

Além disso, uma abordagem inovadora conhecida como Redes de Cápsulas (Capsule Networks ou CapsNet) surgiu como uma alternativa poderosa. As CapsNets utilizam "cápsulas" (grupos de neurônios) para representar propriedades de uma entidade, incluindo sua pose e orientação. Em vez do pooling, empregam um mecanismo de "roteamento por acordo" que preserva as relações espaciais hierárquicas. Essa propriedade confere maior robustez a rotações e translações (variância), e estudos relatam que as CapsNets podem superar as CNNs tradicionais, especialmente em cenários com amostras de treinamento menores (KRUTHIKA; RAJESWARI; MAHESHAPPA, 2019).

2.4.3 Aprendizagem por Transferência

O treinamento de redes neurais profundas a partir do zero é um processo que demanda um volume massivo de dados rotulados, um requisito frequentemente inviável em domínios especializados como a medicina. Para contornar essa limitação, a Aprendizagem por Transferência (*transfer learning*) estabeleceu-se como uma estratégia fundamental (PAN; YANG, 2010; ZHUANG et al., 2021). A abordagem consiste em aproveitar o conhecimento de um modelo previamente treinado em um grande conjunto de dados de

propósito geral, como o ImageNet (DENG et al., 2009), que contém milhões de imagens e ensina à rede a reconhecer características hierárquicas genéricas, como bordas, texturas e formas.

Na prática, a técnica pode ser implementada de duas formas principais (YOSINSKI et al., 2014). A primeira envolve utilizar o modelo pré-treinado como um extrator de características fixo, congelando os pesos das camadas convolucionais e treinando apenas as camadas densas finais (*fully connected layers*), que atuam como o classificador específico para a nova tarefa. Uma abordagem mais poderosa, conhecida como ajuste fino (*fine-tuning*), consiste em reotimizar também os pesos dos blocos de camadas superiores do modelo pré-treinado, permitindo que as características mais abstratas se adaptem ao novo domínio de dados (RAGHU et al., 2019). Essa estratégia não só acelera a convergência do treinamento, mas frequentemente melhora o desempenho final e a capacidade de generalização do modelo, especialmente quando há compatibilidade entre os domínios de origem e destino.

3 Trabalhos Relacionados

A crescente prevalência da Doença de Alzheimer (DA) impulsionou a pesquisa de métodos diagnósticos automatizados, movendo o foco dos métodos estatísticos tradicionais para as técnicas de Aprendizado Profundo. A análise de neuroimagens, como a Ressonância Magnética (MRI), tornou-se o principal campo de aplicação para a Inteligência Artificial (IA) no diagnóstico de DA. Este capítulo apresenta uma revisão da literatura, comparando as principais arquiteturas e técnicas empregadas, e identificando as lacunas que motivam o presente trabalho.

A detecção automatizada da DA migrou de métodos tradicionais para o Aprendizado Profundo, impulsionado pela capacidade das Redes Neurais Convolucionais (CNNs) de extrair características complexas diretamente das neuroimagens. Trabalhos como o de Salehi et al. (2020) e El-Geneedy et al. (2023) confirmaram a superioridade das CNNs, propondo arquiteturas que atingiram alta acurácia na classificação da DA. Para lidar com a natureza tridimensional dos exames, Feng et al. (2019) propuseram uma estrutura híbrida 3D-CNN com uma Rede Neural Recorrente (FSBi-LSTM), demonstrando a importância de capturar a informação espacial 3D. Para superar a escassez de dados médicos, a Aprendizagem por Transferência (*transfer learning*) tornou-se uma estratégia dominante. Lu et al. (2022) utilizaram uma arquitetura Inception-ResNet-V2 com *transfer learning* e alcançaram um feito notável: além de alta acurácia no dataset ADNI, o modelo generalizou com sucesso para bases de dados independentes, incluindo o MIRIAD. No entanto, a eficácia do *transfer learning* em imagens médicas não é universal. O trabalho seminal de Raghu et al. (2019) demonstrou que, em certos cenários, os benefícios de desempenho do pré-treinamento em ImageNet podem ser limitados, sugerindo que as características aprendidas em imagens naturais nem sempre se transferem bem para domínios médicos específicos, o que representa uma lacuna importante na compreensão de quando e como aplicar essa técnica.

Enquanto arquiteturas como ResNet se consolidaram, abordagens inovadoras surgiram. As Redes de Cápsulas (Capsule Networks ou CapsNet) foram propostas para melhor preservar as relações espaciais. Kruthika, Rajeswari e Maheshappa (2019) demonstraram que a CapsNet pode superar uma 3D-CNN, especialmente com amostras de treinamento menores. Apesar desse potencial, a literatura sobre a aplicação de CapsNets para a classificação de Alzheimer ainda é escassa. A análise da literatura aponta para lacunas claras: a generalização dos modelos para datasets externos continua sendo o maior desafio; a interpretabilidade dos modelos de "caixa-preta" é uma barreira para a confiança clínica; e falta uma comparação sistemática de arquiteturas inovadoras, como Vision Transformers

(ViT) e CapsNets, em um mesmo ambiente experimental. Este trabalho se propõe a endereçar esta última lacuna, fornecendo um comparativo robusto em múltiplas bases de dados.

A análise da literatura aponta para lacunas claras: a generalização dos modelos para *datasets* externos continua sendo o maior desafio; a interpretabilidade dos modelos de “caixa-preta” é uma barreira para a confiança clínica; e falta uma comparação sistemática de arquiteturas de vanguarda, como *Vision Transformers* (ViT) e *CapsNets*, em um mesmo ambiente experimental.

Diversos estudos recentes têm demonstrado o potencial das CNNs para classificação de DA a partir de imagens de MRI. El-Assy et al. (2024) propuseram uma arquitetura CNN customizada que alcançou 99,48% de acurácia na detecção precoce de DA utilizando dados do **ADNI**. Shamrat et al. (2023) desenvolveram a AlzheimerNet, um modelo baseado em mudanças funcionais cerebrais que obteve 98,3% de acurácia multi-classe também avaliado no *dataset* **ADNI**. Na linha de arquiteturas híbridas, Vinukonda et al. (2025) integraram DeepLabV3+ para segmentação com LeNet-5 e EResNext, alcançando 98,12% de acurácia e AUC de 0,97 no mesmo banco de dados.

Modelos baseados em *transfer learning* também têm mostrado resultados promissores. Lu et al. (2022) treinaram um classificador prático com 85.721 amostras provenientes de múltiplas fontes clínicas, incluindo **ADNI**, **AIBL**, **OASIS** e **MIRIAD**, demonstrando a importância de grandes volumes de dados heterogêneos. El-Geneedy et al. (2023) utilizaram modelos pré-treinados com *fine-tuning* específico para MRI de DA no **ADNI**, enquanto Alruily et al. (2025) propuseram um *ensemble* de VGG16, MobileNet e InceptionResNetV2 que atingiu 97,93% de acurácia na classificação binária da base **OASIS**.

Recentemente, abordagens baseadas em atenção têm emergido. Saoud e AlMarzouqi (2024) desenvolveram um *ensemble* de 138 *Vision Transformers* 3D focados em ROIs específicas do **ADNI**, utilizando *Deep Belief Networks* para agregação. Esta abordagem híbrida aproveitou tanto a capacidade dos *transformers* de capturar dependências de longo alcance quanto a eficácia de *ensembles* para robustez.

Este trabalho se propõe a endereçar esta última lacuna, fornecendo um comparativo robusto de arquiteturas clássicas (CNNs) e modernas (ViT, CapsNet) em múltiplas bases de dados (MIRIAD e ADNI), avaliando sistematicamente seu desempenho com e sem estratégias de otimização avançadas.

4 Metodologia

Este capítulo descreve em detalhe os procedimentos metodológicos adotados para o desenvolvimento e a avaliação dos modelos de diagnóstico da Doença de Alzheimer. A metodologia foi desenhada para garantir a reprodutibilidade, a robustez e a validade estatística dos resultados. As seções a seguir abordam a base de dados utilizada, as etapas de pré-processamento das imagens, a arquitetura dos modelos implementados, o ambiente de desenvolvimento e as métricas de avaliação de desempenho.

4.1 Bases de Dados

Para este trabalho, foram utilizados dois conjuntos de dados de neuroimagem públicos e amplamente reconhecidos na comunidade de pesquisa: MIRIAD e ADNI. A utilização de múltiplos *datasets* visa fortalecer a validação dos modelos e avaliar sua capacidade de generalização.

4.1.1 MIRIAD

O MIRIAD (*Minimal Interval Resonance Imaging in Alzheimer's Disease*) é um conjunto de dados longitudinal público focado em imagens de ressonância magnética estrutural (sMRI) de pacientes com Doença de Alzheimer e controles cognitivamente saudáveis (MALONE et al., 2013). Desenvolvido pelo Dementia Research Centre do University College London, o MIRIAD tornou-se um importante *benchmark* para validação de algoritmos de diagnóstico automatizado de DA, sendo amplamente utilizado pela comunidade científica devido à sua disponibilidade pública e qualidade dos dados (LU et al., 2022; PEREZ et al., 2023).

O *dataset* contém um total de 708 volumes de MRI estrutural (sequências T1-weighted) provenientes de 69 indivíduos, sendo 46 pacientes diagnosticados com Alzheimer (AD) em estágio leve a moderado e 23 controles saudáveis (*Healthy Control* – HC) pareados por idade. As imagens foram adquiridas em um scanner Signa 1.5T (General Electric). É importante notar que equipamentos de 1.5 Tesla possuem menor intensidade de campo magnético em comparação aos scanners modernos de 3T ou superiores, o que pode resultar em imagens com menor relação sinal-ruído e resolução de contraste, impondo um desafio adicional aos algoritmos de classificação. As imagens possuem resolução de $256 \times 256 \times 124$ voxels e tamanho de voxel de $0.9375 \times 0.9375 \times 1.5$ mm³, utilizando uma sequência SPGR (*Spoiled Gradient Echo*) (MALONE et al., 2013).

A principal característica distintiva do MIRIAD é a coleta de imagens em intervalos mínimos de tempo, variando de 2 semanas a 2 anos entre as sessões de imageamento, o que o torna particularmente valioso para estudos longitudinais sobre a progressão da doença e a taxa de atrofia cerebral ao longo do tempo (SILVA et al., 2023). Cada participante foi submetido a múltiplas sessões de MRI (média de 10 scans por indivíduo com AD e 11 scans por controle), permitindo análise temporal detalhada da neurodegeneração.

Para este trabalho, o MIRIAD foi utilizado em um contexto transversal, considerando apenas uma imagem por paciente (tipicamente a sessão *baseline*), resultando em uma tarefa de classificação binária: 46 amostras da classe AD versus 23 amostras da classe HC. Esta configuração apresenta um desbalanceamento de classes com razão aproximada de 2:1 (AD:HC), o que exigiu estratégias específicas de tratamento durante o treinamento, como técnicas de *data augmentation*, balanceamento por pesos de classe e validação cruzada estratificada para garantir representatividade adequada da classe minoritária (PATIL et al., 2022).

O diagnóstico de DA no MIRIAD foi confirmado por avaliação clínica seguindo critérios NINCDS-ADRDA, e os controles saudáveis foram cuidadosamente selecionados para não apresentarem sinais de comprometimento cognitivo ou anormalidades estruturais cerebrais. O *dataset* também fornece dados demográficos (idade, sexo), escores do Mini-Exame do Estado Mental (MEEM) e volumes de estruturas cerebrais segmentadas automaticamente, embora neste trabalho o foco tenha sido exclusivamente nas imagens brutas de MRI para classificação automatizada.

O MIRIAD tem sido utilizado como *benchmark* de validação externa em diversos estudos recentes de *deep learning* para DA, com trabalhos reportando acurácias superiores a 99% utilizando CNNs 3D (PATIL et al., 2022) e 89,1% de sensibilidade com abordagens estatísticas (PEREZ et al., 2023), demonstrando ser um *dataset* desafiador mas adequado para avaliação de generalização de modelos treinados em outras bases como ADNI.

4.1.2 ADNI

A Alzheimer’s Disease Neuroimaging Initiative (ADNI) é uma das iniciativas mais influentes e extensas para a coleta de dados sobre a Doença de Alzheimer, sendo considerada o padrão-ouro para desenvolvimento e validação de algoritmos diagnósticos automatizados (JACK et al., 2018; Alzheimer’s Disease Neuroimaging Initiative (ADNI), 2024). Lançada em 2004 como um estudo colaborativo multicêntrico envolvendo mais de 50 instituições nos Estados Unidos e Canadá, a ADNI tem como objetivo principal validar biomarcadores para ensaios clínicos de prevenção e tratamento da DA, além de fornecer dados padronizados de alta qualidade para a comunidade científica internacional (PETERSEN et al., 2010).

O projeto ADNI foi estruturado em múltiplas fases sequenciais:

- **ADNI-1 (2004-2010):** recrutou 819 participantes iniciais, sendo 229 controles cognitivamente normais (CN, *Cognitively Normal*), 398 com comprometimento cognitivo leve (MCI, *Mild Cognitive Impairment*) e 192 com Alzheimer em estágio leve (AD, *Alzheimer's Disease*) (PETERSEN et al., 2010). Os participantes foram acompanhados longitudinalmente por até 36 meses.
- **ADNI-GO (2009-2011):** expandiu o estudo adicionando 200 novos participantes com comprometimento cognitivo leve precoce (EMCI, *Early Mild Cognitive Impairment*), representando um estágio ainda mais inicial da doença.
- **ADNI-2 (2011-2016):** recrutou 550 novos participantes distribuídos nas categorias CN, EMCI, MCI tardio (LMCI, *Late Mild Cognitive Impairment*) e AD, além de introduzir uma nova categoria de Preocupação Significativa com Memória (SMC, *Significant Memory Concern*) com 107 participantes (NYU Langone Health, 2025).
- **ADNI-3 (2016-presente):** continua o acompanhamento longitudinal e incorpora novas tecnologias de imageamento, incluindo PET com marcadores tau (Flortaucipir).

Até o momento, a ADNI acumulou dados de mais de 2.400 participantes ao longo de suas fases, com mais de 15.000 volumes de imagens de ressonância magnética estrutural (T1-weighted) disponíveis (ADNI Clinical Core, 2024). As imagens de MRI foram adquiridas em scanners seguindo protocolos rigorosamente padronizados, com resoluções típicas de $256 \times 256 \times 170$ voxels e tamanho de voxel variando entre $1.0 \times 1.0 \times 1.2 \text{ mm}^3$ (1.5T) e $1.0 \times 1.0 \times 1.0 \text{ mm}^3$ (3.0T). O pré-processamento padrão inclui correção de não-uniformidade de campo magnético, registro espacial ao template MNI152 e segmentação automatizada de tecidos e estruturas cerebrais.

A ADNI oferece um espectro diagnóstico detalhado baseado em avaliações clínicas padronizadas e testes neuropsicológicos, permitindo classificações desde multiclasse (CN/EMCI/LMCI/AD) até binária (AD vs CN), além de tarefas de predição de conversão de MCI para AD.

Para este trabalho, foram utilizadas as imagens de MRI estrutural T1-weighted do ADNI, selecionando um subconjunto de dados de baseline (visita inicial) para uma tarefa de classificação multiclasse em 4 categorias. A Tabela 1 apresenta a distribuição detalhada das amostras utilizadas por classe diagnóstica neste estudo.

O dataset apresenta uma progressão natural dos estágios da doença: desde indivíduos sem demência (**Cognitively Normal** — **CN**), passando por demência muito leve (**Early Mild Cognitive Impairment** — **EMCI**, correspondente ao estágio MCI

Tabela 1 – Distribuição de Amostras (Fatias 2D) do Dataset ADNI por Classe

Classe	Treino	Teste	Total	(%)
VeryMildDementia	2566	634	3200	50.0%
ModerateDementia	1781	459	2240	35.0%
MildDementia	724	172	896	14.0%
NonDemented	49	15	64	1.0%
Total	5120	1280	6400	100.0%

Fonte: O autor (2025). Total de fatias 2D extraídas de aproximadamente 2.133 volumes (3 fatias/volume).
Divisão por paciente: 80% treino/validação, 20% teste.

inicial), demência leve (**Late Mild Cognitive Impairment** — **LMCI**, estágio MCI avançado ou AD inicial) até demência moderada (**Alzheimer’s Disease** — **AD**, AD estabelecida). Esta configuração multiclasse permite ao modelo aprender a distinguir não apenas entre presença/ausência da doença, mas também o grau de severidade do comprometimento cognitivo, tornando a tarefa significativamente mais desafiadora do que a classificação binária tradicional.

A classificação multiclasse com 4 categorias apresenta desafios únicos em relação à abordagem binária tradicional (AD vs CN):

- **Maior complexidade de aprendizado:** o modelo precisa aprender fronteiras de decisão entre quatro classes, não apenas duas.
- **Desbalanceamento extremo:** a razão entre a classe majoritária (**EMCI**) e a minoritária (**CN**) é de 50:1, exigindo tratamento especializado.
- **Confusão entre classes adjacentes:** classes contíguas no espectro da doença (ex.: **EMCI** vs **LMCI**) apresentam sobreposição de características clínicas e de imagem.
- **Métricas mais exigentes:** acurácia isolada pode ser enganosa; métricas balanceadas como F1-Score macro e MCC são essenciais.

Um desafio crítico deste estudo no ADNI é o desbalanceamento extremo de classes observado no subconjunto utilizado. A classe **EMCI** representa 50% das amostras (3.200 imagens), enquanto a classe minoritária **CN** representa apenas 1% do total (64 imagens), resultando em uma razão de desbalanceamento de 50:1 entre a classe majoritária e a minoritária. Este desbalanceamento severo exigiu estratégias rigorosas de tratamento, incluindo: (i) pesos de classe (*class weights*) calculados inversamente proporcionais à frequência de cada categoria, (ii) técnicas avançadas de *data augmentation* aplicadas de forma mais agressiva às classes minoritárias, (iii) validação cruzada estratificada para

garantir representatividade proporcional de todas as classes em cada fold, e (iv) monitoramento de métricas balanceadas (F1-Score Macro e MCC) em vez de acurácia simples, para evitar viés em direção à classe majoritária.

A ADNI tornou-se referência mundial em pesquisas sobre DA, com mais de 2.000 publicações científicas baseadas em seus dados. Estudos recentes de *deep learning* reportam acurácias superiores a 95% para classificação binária (AD vs CN) (LU et al., 2022; EL-GENEEDY et al., 2023) e acima de 90% para classificação multiclasse (SHAMRAT et al., 2023; VINUKONDA et al., 2025), consolidando a base como benchmark essencial para validação de algoritmos de IA aplicados ao diagnóstico de Alzheimer.

4.2 Pré-processamento

Um *pipeline* de pré-processamento rigoroso foi implementado para padronizar as imagens e prepará-las para o treinamento dos modelos de *Deep Learning*. A Figura 5 apresenta exemplos de fatias axiais do *dataset* MIRIAD, ilustrando a variabilidade dos dados de entrada utilizados neste estudo.

É importante destacar uma particularidade na estruturação dos dados de entrada para as redes pré-treinadas (como ResNet e EfficientNet), que originalmente esperam imagens coloridas compostas por três canais (RGB). Neste trabalho, a entrada foi adaptada utilizando três fatias axiais consecutivas do cérebro (uma fatia central e duas adjacentes) para preencher esses três canais. Embora essa abordagem permita o aproveitamento da arquitetura original sem modificar a camada de entrada, ela introduz uma limitação potencial no processo de *Transfer Learning*: os filtros das primeiras camadas, treinados para detectar padrões de cor e bordas em imagens naturais, precisam se readaptar para interpretar variações de intensidade de cinza volumétricas como se fossem informações espectrais. Essa adaptação pode exigir um ajuste mais fino dos pesos durante o treinamento para que a rede aprenda a correlacionar a continuidade espacial entre as fatias com a representação interna de canais.

4.2.1 Normalização e Padronização de Intensidade

O processo inicia com o carregamento do arquivo NIfTI 3D. A partir do volume tridimensional, foram extraídas múltiplas fatias bidimensionais (2D) do plano axial. Especificamente, para aumentar a quantidade de dados de treinamento sem recorrer a múltiplas reutilizações dos dados do mesmo paciente, foram selecionadas três fatias axiais de cada volume: uma fatia central (corte médio do cérebro) e duas fatias adjacentes ($\pm 10\text{mm}$ da posição central), totalizando aproximadamente 3 fatias por volume cerebral.

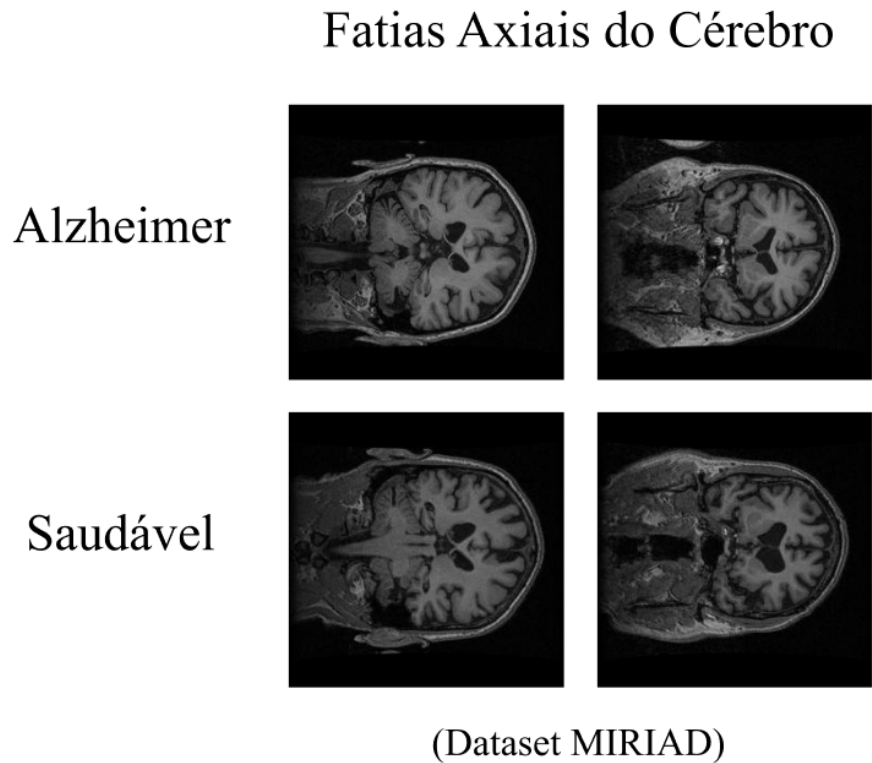


Figura 5 – Exemplos de Fatias Axiais do Dataset MIRIAD

Fonte: O autor (2025).

Esta abordagem de extração multi-slice representa uma estratégia de aumento de dados (*data augmentation*) controlada que preserva a independência entre pacientes, pois todas as fatias de um mesmo paciente permanecem juntas durante a divisão treino/teste. Embora esta escolha descarte parte da informação volumétrica tridimensional completa, ela permite: (i) triplicar o tamanho do conjunto de dados mantendo a separação estrita por paciente, (ii) capturar variabilidade anatômica em diferentes níveis do eixo vertical do cérebro, e (iii) aproveitar modelos de Aprendizagem por Transferência pré-treinados em bases 2D como ImageNet.

4.2.2 Redimensionamento

Todas as imagens, após a normalização, foram redimensionadas para uma resolução fixa de 224x224 pixels. Este tamanho é um padrão para muitas arquiteturas de *Deep Learning* modernas, incluindo ResNet, EfficientNet e Vision Transformers, permitindo o aproveitamento dos pesos pré-treinados.

4.2.3 Aumento de Dados (Data Augmentation)

Para aumentar a diversidade do conjunto de treinamento e reduzir o risco de sobreajuste (*overfitting*), um conjunto de técnicas de aumento de dados foi aplicado em tempo de execução, apenas nas imagens de treino. As transformações incluíram:

- **Transformações Geométricas:** Rotações aleatórias (± 10 graus), espelhamento horizontal (com 30% de probabilidade) e translações afins ($\pm 5\%$ da dimensão da imagem). Neste contexto, a translação afim refere-se a um deslocamento linear dos pixels no plano x, y , preservando o paralelismo das linhas, utilizada para simular pequenas variações no posicionamento da cabeça do paciente dentro do campo de visão.
- **Transformações de Cor e Intensidade:** Variações aleatórias de brilho e contraste ($\pm 10\%$).
- **Transformações de Textura:** Aplicação de desfoque gaussiano (*Gaussian Blur*) com um *kernel* de tamanho 3×3 e sigma variável.

O pipeline completo de pré-processamento e treinamento, desde a entrada dos dados brutos até a geração das previsões finais, é ilustrado de forma esquemática na Figura 6.

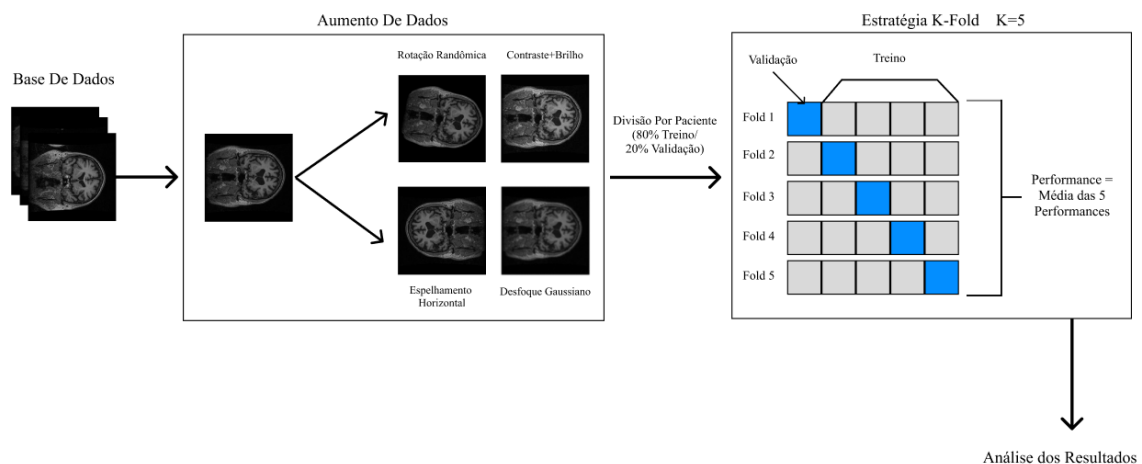


Figura 6 – Pipeline Completo de Pré-processamento e Treinamento dos Dados

Fonte: O autor (2025).

4.2.4 Divisão dos Dados e Validação Cruzada

Para garantir uma avaliação robusta e evitar o vazamento de dados (*data leakage*), a divisão dos dados foi realizada por paciente. Primeiramente, 20% dos pacientes foram separados para formar um conjunto de teste final (*holdout*), que não foi utilizado durante nenhuma etapa de treinamento ou validação. Os 80% restantes dos pacientes foram utilizados em um esquema de Validação Cruzada K-Fold Estratificada por Grupos, com $k = 5$, conforme ilustrado na Figura 6.

Na abordagem ora adotada, os dados foram divididos em 5 folds (subconjuntos), garantindo que todos os exames de um mesmo paciente pertencessem a um único fold. A cada iteração, um fold foi usado para validação e os outros quatro para treinamento. A estratificação garantiu que a proporção entre as classes fosse mantida em cada fold: para MIRIAD, as classes AD e HC; para ADNI, as quatro categorias diagnósticas (**Cognitively Normal — CN**, **Early Mild Cognitive Impairment — EMCI**, **Late Mild Cognitive Impairment — LMCI** e **Alzheimer’s Disease — AD**). Esta técnica permite uma avaliação mais confiável da capacidade de generalização do modelo.

4.3 Arquiteturas dos Modelos e Estratégia de Treinamento

Foram implementadas e comparadas seis arquiteturas de *Deep Learning* distintas, abrangendo diferentes paradigmas de visão computacional. O fluxo de processamento e a estratégia de otimização foram padronizados para garantir uma comparação justa entre os modelos.

O fluxo do sistema inicia-se com o carregamento do arquivo NIFTI 3D, seguido pela extração da fatia 2D axial. Esta imagem passa pelo pipeline de pré-processamento e, durante o treino, pelas transformações de aumento de dados. O tensor resultante é então alimentado ao modelo, que realiza a inferência. A perda é calculada, os pesos são atualizados pelo otimizador e, por fim, as métricas de desempenho são registradas para análise.

4.3.1 Detalhamento das Arquiteturas Implementadas

As seis arquiteturas avaliadas foram:

- **CNN Manual:** Uma Rede Neural Convolutiva customizada, composta por três blocos sequenciais de convolução (com 32, 64 e 128 filtros respectivamente), cada um seguido por Batch Normalization, ativação ReLU, Max Pooling e Dropout (com p de 0.25 a 0.3). O classificador final utiliza uma camada de Adaptive Average

Pooling, seguida por camadas lineares (512 e 128 neurônios) com regularização por Dropout ($p=0.5$ e $p=0.3$) antes da camada de saída final.

- **Vision Transformer (ViT-BT):** Utilizou-se o modelo *google/vit-base-patch16-224-in21k* como base. A cabeça de classificação original foi substituída por uma única camada linear (`nn.Linear`) que recebe o vetor de 768 dimensões do token de classificação ([CLS]) e o projeta para as 2 ou 4 saídas da tarefa.
- **HuggingFace (ViT Fine-Tuned):** Um modelo ViT similar ao anterior, mas que já havia passado por um processo de *fine-tuning* para uma tarefa de visão computacional correlata, obtido da plataforma Hugging Face. Sua cabeça de classificação também foi adaptada para 2 ou 4 saídas.
- **ResNet152:** Uma rede residual com 152 camadas, pré-treinada no dataset ImageNet. A camada classificadora final (`fc`) foi substituída por uma nova camada `nn.Linear` que recebe as 2048 características do último bloco residual e as projeta para as 2 ou 4 saídas.
- **EfficientNet-V2-M:** Uma arquitetura moderna e eficiente, pré-treinada no ImageNet. O classificador final foi substituído por uma nova camada `nn.Linear` que recebe as 1280 características do último estágio convolucional e as mapeia para 2 ou 4 saídas.
- **CapsNet:** Uma Rede de Cápsulas customizada, consistindo em: (1) uma camada convolucional inicial com 256 filtros (kernel 9x9, stride 4); (2) uma camada de cápsulas primárias (*PrimaryCaps*) gerando aproximadamente 4.232 cápsulas de 8 dimensões; e (3) uma camada final de cápsulas de dígitos (DigitCaps) com 2 ou 4 cápsulas de 16 dimensões (dependendo do dataset: 2 para MIRIAD binário, 4 para ADNI multiclasse). O modelo utiliza 3 iterações de roteamento dinâmico.

4.3.2 Configuração das Camadas de Saída

Para garantir compatibilidade com as diferentes tarefas de classificação de cada dataset, as arquiteturas foram configuradas com camadas de saída específicas:

- Para o dataset MIRIAD (tarefa binária AD vs HC): todas as arquiteturas foram configuradas com camadas finais projetando para 2 saídas.
- Para o dataset ADNI (tarefa multiclasse de 4 categorias): todas as arquiteturas foram configuradas com camadas finais projetando para 4 saídas correspondentes às classes **CN**, **EMCI**, **LMCI** e **AD**.

Esta abordagem permite avaliar a capacidade de cada arquitetura tanto em classificação binária quanto em classificação multiclasse desbalanceada, fornecendo insights sobre sua versatilidade e robustez.

4.3.3 Hiperparâmetros e Otimização

Todos os modelos foram treinados com uma estratégia de otimização consistente para garantir uma comparação justa entre as diferentes arquiteturas. A Tabela 2 resume os principais hiperparâmetros e configurações utilizadas na fase de treinamento avançado (Etapa 1).

4.4 Ambiente

Os experimentos foram conduzidos em um servidor com sistema operacional Linux, equipado com uma GPU NVIDIA GeForce RTX 4090. A linguagem de programação foi Python, e o *framework* de *Deep Learning* utilizado foi o PyTorch. A Tabela 2 sumariza as principais bibliotecas auxiliares. O código completo e as rotinas de experimentação estão disponíveis em um repositório GitHub para garantir a reprodutibilidade. O gerenciamento e o rastreamento dos experimentos foram realizados com a ferramenta MLflow.

Tabela 2 – Hiperparâmetros e Configurações de Otimização.

Parâmetro	Valor / Configuração
Otimizador	Adamax
Taxa de Aprendizagem Inicial	0.001 (CNN, CapsNet) / 0.0001 (Modelos pré-treinados)
Tamanho do Lote (Batch Size)	32 (Padrão) / 16 (CapsNet)
Função de Perda	Cross-Entropy Loss / Margin Loss (CapsNet)
Agendador de Taxa de Aprendizagem	ReduceLROnPlateau
- Fator de Redução	0.5
- Paciência (Scheduler)	3 épocas
Parada Antecipada (Early Stopping)	Sim
- Métrica Monitorada	Perda de Validação (val_loss)
- Paciência (Early Stopping)	10 épocas
Regularização Adicional	Pesos de Classe (Class Weights) para desbalanceamento

4.5 Métricas de Avaliação

Para avaliar e comparar o desempenho dos modelos de forma quantitativa e robusta em uma tarefa de classificação multiclasse (4 classes), foi utilizado um conjunto abrangente de métricas.

Tabela 3 – Bibliotecas Python Auxiliares

Biblioteca	Finalidade Principal
NumPy/Pandas	Manipulação de dados numéricos e metadados.
Scikit-learn	Divisão de dados, validação cruzada e cálculo de métricas.
Nibabel	Leitura de arquivos de neuroimagem no formato NIfTI.
Pillow (PIL)	Manipulação e transformação de imagens.
Matplotlib/Seaborn	Geração de gráficos e visualização de dados.
MLflow	Rastreamento e gerenciamento de experimentos.

Fonte: O autor (2025).

A matriz de confusão 4×4 é a base para a maioria das métricas, tabulando as predições corretas e incorretas para cada uma das 4 classes, permitindo uma análise detalhada dos padrões de confusão entre categorias adjacentes no espectro da doença.

A acurácia mede a proporção geral de acertos sobre todas as classes. Para problemas multiclasse desbalanceados, esta métrica pode ser enganosa, pois um modelo que sempre prediz a classe majoritária pode alcançar acurácia artificialmente alta.

O F1-Score Macro calcula o F1-Score para cada classe separadamente e então computa a média não ponderada, tratando todas as classes igualmente independente de seu tamanho. Esta métrica é crucial para avaliar o desempenho balanceado em datasets desbalanceados.

O F1-Score Ponderado (Weighted) computa a média dos F1-Scores de cada classe ponderada pela frequência de cada classe no conjunto de dados, fornecendo uma métrica que reflete melhor o desempenho geral considerando a distribuição real.

O Matthews Correlation Coefficient (MCC) pode ser estendido para multiclasse e fornece uma medida balanceada que considera verdadeiros/falsos positivos/negativos de todas as classes, sendo particularmente robusto a desbalanceamento extremo. Seus valores variam de -1 (predição completamente errada) a +1 (predição perfeita).

Para classificação multiclasse, também reportamos Precisão, Revocação (Sensibilidade) e Especificidade por classe, calculadas usando abordagem One-vs-Rest (cada classe contra todas as outras).

5 Resultados

Este capítulo apresenta e analisa os resultados experimentais obtidos a partir da metodologia descrita no Capítulo 4. O objetivo é avaliar o desempenho das diferentes arquiteturas de *Deep Learning* e estratégias de otimização na classificação da Doença de Alzheimer, utilizando os datasets MIRIAD e ADNI.

A análise é estruturada em três estágios progressivos de complexidade, seguida por uma discussão crítica das limitações do trabalho e uma comparação com a literatura.

5.1 Análise Exploratória dos Dados

Os experimentos foram conduzidos utilizando dois datasets públicos: MIRIAD e ADNI. O dataset MIRIAD serviu como a base principal para o desenvolvimento e comparações iniciais, enquanto o dataset ADNI foi utilizado para avaliar a capacidade de generalização e a eficácia das estratégias de otimização implementadas em nosso estudo.

A distribuição de classes foi verificada em ambos os datasets: classificação binária (AD vs HC) no MIRIAD e classificação multiclasse (4 categorias de severidade) no ADNI, e o desbalanceamento inerente foi tratado por meio de pesos de classe (class weights) durante o treinamento, conforme descrito no capítulo de Metodologia. Conforme previamente descrito, a divisão dos dados foi realizada estritamente por paciente, com 80% dos indivíduos alocados para o conjunto de treinamento e validação (utilizado na validação cruzada de 5 folds) e 20% para o conjunto de teste final.

5.2 Análise de Desempenho por Estágio

Os resultados são apresentados em três seções, correspondentes às três etapas experimentais: a linha de base (Etapa 0), o treinamento com otimizações avançadas (Etapa 1) e as estratégias de ajuste fino (fine-tuning) (Etapa 2). A análise se concentrará nas métricas de Acurácia, Sensibilidade, Especificidade, MCC e AUC-ROC para avaliar o desempenho dos modelos nos datasets MIRIAD e ADNI.

5.2.1 Etapa 0: Linha de Base

A primeira etapa consistiu em um treinamento utilizando modelos pré-treinados no ImageNet com *fine-tuning* completo de todas as camadas, sem o emprego de estratégias avançadas de otimização como aumento de dados ou validação cruzada.

Os resultados das métricas de avaliação no conjunto de teste final de cada dataset são apresentados nas Tabelas 4 e 5.

Tabela 4 – Resultados da Etapa 0 (Linha de Base) no Dataset **MIRIAD**

Modelo	Acurácia	Sensibilidade	Especificidade	MCC	AUC-ROC
CNN Manual	0.886	0.886	0.978	0.760	0.954
ResNet152	0.805	0.805	0.763	0.619	0.911
HuggingFace	0.765	0.765	0.925	0.485	0.828
EfficientNet-V2-M	0.752	0.752	0.860	0.455	0.829
CapsNet	0.691	0.691	0.892	0.301	0.769
ViT-BT	0.658	0.658	0.903	0.205	0.727

Fonte: O autor (2025). Resultados no conjunto de teste final. Melhores valores em negrito.

No dataset MIRIAD, a CNN Manual customizada obteve o melhor desempenho global, alcançando 88.6% de acurácia. Esse resultado sugere que, em cenários com volume reduzido de dados, arquiteturas menores e mais específicas para o domínio podem generalizar melhor do que modelos mais profundos, uma vez que exigem menos parâmetros a serem ajustados durante o *fine-tuning*.

Tabela 5 – Resultados da Etapa 0 (Linha de Base) no Dataset **ADNI**

Modelo	Acurácia	Especificidade	AUC-ROC
CapsNet	0.206	0.961	0.484
CNN Manual	0.559	0.817	0.721
EfficientNet-V2-M	0.984	0.990	0.997
HuggingFace	0.572	0.823	0.732
ResNet152	0.975	0.987	0.998
ViT-BT	0.911	0.960	0.974

Fonte: O autor (2025). Resultados no conjunto de teste final. Melhores valores em negrito.

No dataset ADNI, o cenário se inverteu substancialmente. Na tarefa de classificação multiclasse de 4 categorias, as arquiteturas profundas EfficientNet-V2-M e ResNet152 demonstraram desempenho amplamente superior, alcançando acurácias de 98.4% e 97.5%, respectivamente. Em contraste, os modelos customizados (CNN Manual e CapsNet) apresentaram resultados consideravelmente inferiores. Esse comportamento indica que, para um conjunto de imagens mais complexo e com maior variabilidade estrutural, arquiteturas mais profundas e com maior capacidade de extração hierárquica de características conseguem aprender representações mais discriminativas, quando submetidas a *fine-tuning* completo sem otimizações adicionais. Em outras palavras, a complexidade intrínseca da base de imagens ADNI favorece arquiteturas profundas pré-treinadas e mecanismos de regularização mais sofisticados, enquanto arquiteturas mais simples não dispõem de ca-

pacidade suficiente para capturar as variações sutilmente presentes nos volumes cerebrais deste dataset.

5.2.2 Etapa 1: Otimização de Parâmetros e Técnicas de Ajuste

Na segunda etapa, o mesmo conjunto de modelos da Etapa 0 foi submetido a um pipeline de treinamento robusto. Esta fase incorporou um conjunto de otimizações para aprimorar a capacidade de generalização e a estabilidade do treinamento, incluindo: aumento de dados (*data augmentation*), parada antecipada (*early stopping*) para prevenir o sobreajuste, um agendador de taxa de aprendizado (*ReduceLROnPlateau*) para refinar a convergência, e o uso de pesos de classe (*class weights*) para lidar com o desbalanceamento dos dados. O desempenho foi avaliado de forma mais rigorosa, utilizando validação cruzada de 5 folds para obter uma estimativa mais confiável da performance de cada arquitetura. As tabelas a seguir apresentam a média dos resultados obtidos nesses 5 folds

Tabela 6 – Resultados da Etapa 1 (Otimização Avançada) no Dataset **MIRIAD**

Modelo	Acurácia	Sensibilidade	Especificidade	MCC	AUC-ROC
CapsNet	0.698	0.698	0.903	0.112	0.647
CNN Manual	0.671	0.671	1.000	0.533	
EfficientNet-V2-M	0.804	0.804	0.914	0.621	0.859
HuggingFace	0.703	0.703	0.903	0.183	0.704
ResNet152	0.827	0.827	0.935	0.647	0.902
ViT-BT	0.673	0.673	0.925	0.090	0.658

Fonte: O autor (2025). Métricas são a média da validação cruzada. Especificidade é do teste final. Melhores valores em negrito.

Com a otimização, a ResNet152 se consolidou como o modelo mais performático no MIRIAD, com uma acurácia média de 82.7%. A CNN Manual, por outro lado, teve seu desempenho degradado, sugerindo que as técnicas de regularização e aumento de dados não beneficiaram a arquitetura mais simples da mesma forma que os modelos pré-treinados.

No dataset ADNI, os resultados da otimização foram ainda mais expressivos. EfficientNet-V2-M e ResNet152 alcançaram um desempenho excepcional e estável, com acurácias médias de 97.7% e 96.9%, respectivamente, e valores de AUC-ROC próximos de 1.0, indicando uma capacidade de discriminação quase perfeita. Estes resultados confirmam a eficácia do pipeline de treinamento avançado combinado com *fine-tuning* completo em um dataset robusto.

Tabela 7 – Resultados da Etapa 1 (Otimização Avançada) no Dataset **ADNI**

Modelo	Acurácia	Sensibilidade	Especificidade	MCC	AUC-ROC
CapsNet	0.913	0.913	0.961	0.857	0.982
CNN Manual	0.571	0.571	0.817	0.264	0.737
EfficientNet-V2-M	0.977	0.977	0.990	0.963	0.997
HuggingFace	0.579	0.579	0.823	0.289	0.754
ResNet152	0.969	0.969	0.987	0.950	0.997
ViT-BT	0.900	0.900	0.960	0.836	0.972

Fonte: O autor (2025). Métricas são a média da validação cruzada. Melhores valores em negrito.

5.2.3 Etapa 2: Estratégias de *fine-tuning*

Os dois melhores modelos da Etapa 1 (EfficientNet-V2-M e ResNet152) foram selecionados para a fase de ajuste fino, conduzida em ambos os datasets MIRIAD e ADNI. Quatro estratégias distintas de *fine-tuning* foram avaliadas: Custom FC (substituição das camadas totalmente conectadas finais), Full Fine-Tune (ajuste de todas as camadas da rede), Frozen (congelamento completo do extrator de características), e Top Layers (descongelamento apenas das camadas mais profundas). As Tabelas 8 e 9 apresentam os resultados no conjunto de teste final para cada dataset.

Tabela 8 – Resultados da Etapa 2 (Estratégias de *fine-tuning*) no Dataset ADNI

Modelo	Estratégia	Acurácia	Sensibilidade
EfficientNet-V2-M	Custom FC	0.500	0.500
	Frozen	0.510	0.510
	Top Layers	0.550	0.550
	Full Fine-Tune	0.990	0.990
ResNet152	Custom FC	0.445	0.445
	Frozen	0.454	0.454
	Top Layers	0.616	0.616
	Full Fine-Tune	0.921	0.921

Fonte: O autor (2025). Resultados no conjunto de teste final. Melhores valores em negrito.

5.2.3.1 Análise Comparativa dos Resultados de Fine-Tuning

Os resultados da Etapa 2 revelam que a estratégia de *fine-tuning* completo (*Full Fine-Tune*) emergiu como a abordagem mais eficaz para ambos os *datasets*, conforme ilustrado nas Figuras 7 e 8. É importante ressaltar que a estratégia de *Full Fine-Tune* utilizada nesta etapa é metodologicamente idêntica à empregada na Etapa 1. As variações observadas nos valores de desempenho devem-se à realização de uma nova rodada de treinamento (*new run*) independente, sujeita à estocasticidade inerente à inicialização de

Tabela 9 – Resultados da Etapa 2 (Estratégias de *fine-tuning*) no Dataset **MIRIAD**

Modelo	Estratégia	Acurácia	Sensibilidade	Especificidade	MCC
EfficientNet-V2-M	Custom FC	0.658	0.658	0.903	0.205
	Frozen	0.631	0.631	0.989	0.086
	Full Fine-Tune	0.772	0.772	0.882	0.500
	Top Layers	0.772	0.772	0.828	0.510
ResNet152	Custom FC	0.651	0.651	0.892	0.188
	Frozen	0.617	0.617	0.860	0.097
	Full Fine-Tune	0.772	0.772	0.849	0.505
	Top Layers	0.832	0.832	0.925	0.636

Fonte: O autor (2025). Resultados no conjunto de teste final. Melhores valores em negrito. AUC-ROC não disponível para esta etapa.

pesos e ao processo de otimização estocástica. Ainda assim, a consistência dos resultados elevados valida a hipótese de que, quando adequadamente implementado com técnicas robustas de regularização, o ajuste fino profundo de todas as camadas supera estratégias mais conservadoras de congelamento parcial.

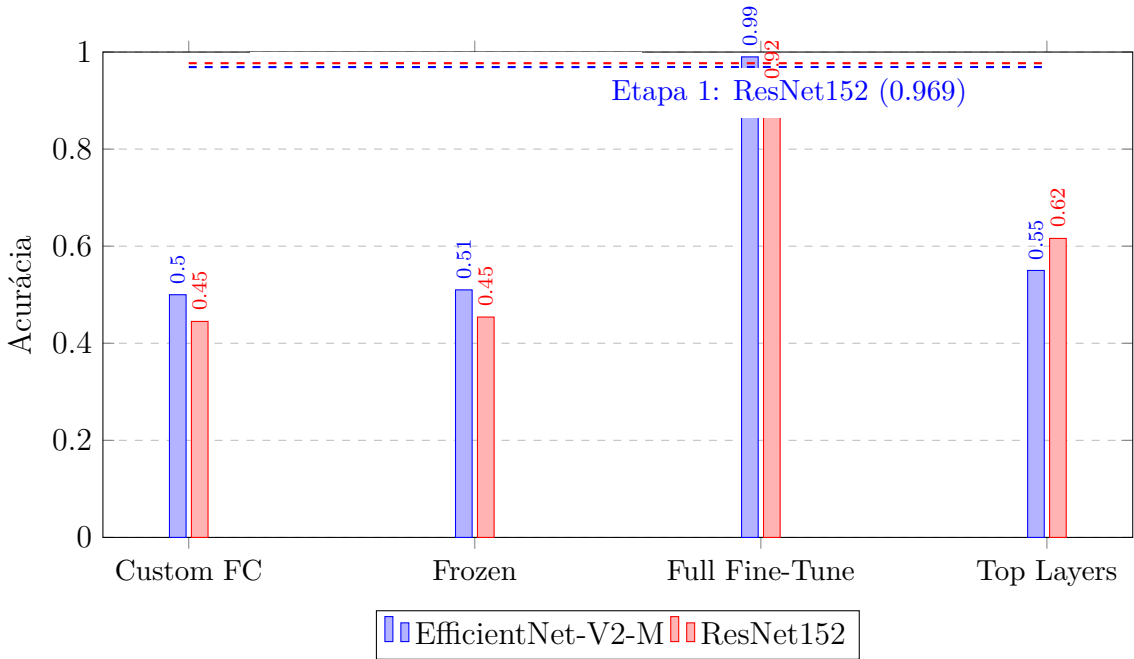


Figura 7 – Desempenho das Estratégias de *fine-tuning* no Dataset ADNI (Etapa 2 - Classificação Multiclasse de 4 Categorias). Linhas tracejadas indicam o desempenho da Etapa 1. A estratégia Full Fine-Tune apresentou desempenho excepcional: 99.0% para EfficientNet-V2-M e 92.1% para ResNet152, demonstrando a eficácia do ajuste completo das camadas mesmo em tarefa multiclasse desbalanceada.

Fonte: O autor (2025).

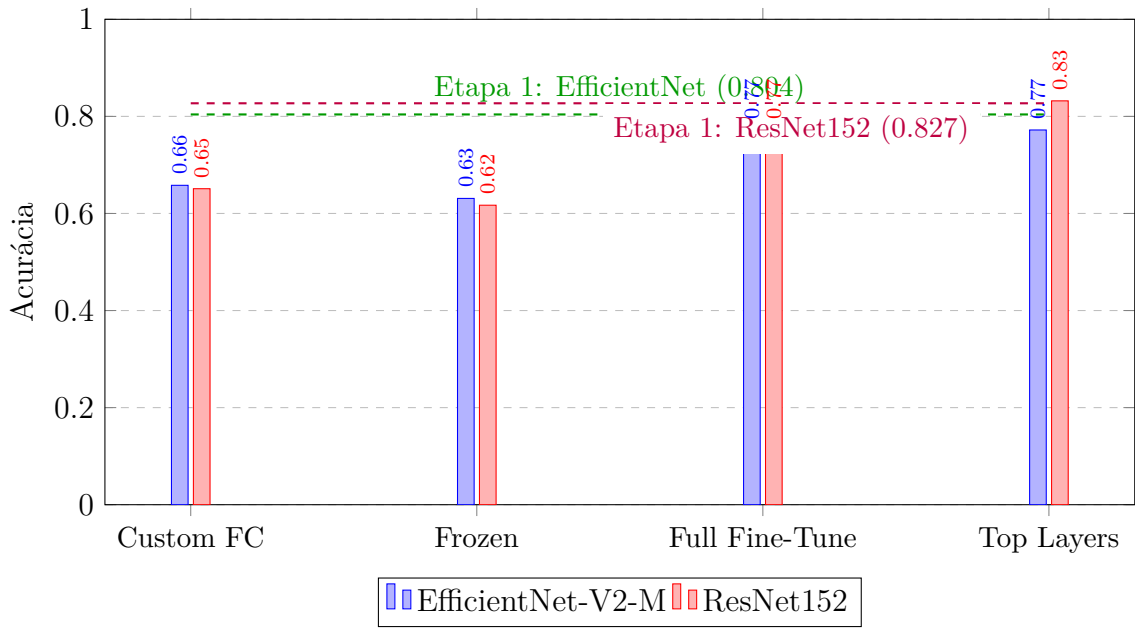


Figura 8 – Desempenho das Estratégias de *fine-tuning* no Dataset MIRIAD (Etapa 2 - Classificação Binária AD vs HC). Linhas tracejadas indicam o desempenho da Etapa 1. A estratégia Top Layers com ResNet152 alcançou o melhor resultado geral (83.2%), superando marginalmente a Etapa 1.

Fonte: O autor (2025).

Resultados no Dataset ADNI: Conforme evidenciado na Figura 7, o desempenho no ADNI apresentou resultados excelentes com a estratégia Full Fine-Tune. A EfficientNet-V2-M alcançou 99.0% de acurácia, enquanto a ResNet152 atingiu 92.1% de acurácia. Embora ligeiramente inferiores aos resultados da Etapa 1 (97.7% e 96.9%, respectivamente), estes valores demonstram que o *fine-tuning* controlado mantém alta capacidade de generalização.

As estratégias conservadoras (Frozen, Custom FC e Top Layers) apresentaram desempenho substancialmente inferior, com acurácias variando entre 44.5% e 61.6%, confirmando que, para o dataset ADNI com sua maior complexidade e variabilidade, o ajuste completo das camadas é necessário para extrair representações discriminativas adequadas.

Resultados no Dataset MIRIAD: O cenário no MIRIAD, ilustrado na Figura 8, apresentou resultados ainda mais promissores. A estratégia Top Layers aplicada à ResNet152 alcançou o melhor desempenho geral com 83.2% de acurácia e MCC de 0.636, superando marginalmente o baseline da Etapa 1 (82.7%). A estratégia Full Fine-Tune também demonstrou forte desempenho, atingindo 77.2% de acurácia com especificidades equilibradas acima de 84.9% para ambos os modelos.

Importante destacar que, em todas as configurações testadas no MIRIAD, a especificidade permaneceu em níveis clinicamente adequados (acima de 82.8%), com a estratégia Frozen da EfficientNet-V2-M alcançando notáveis 98.9% de especificidade, demonstrando

que os modelos preservaram sua capacidade discriminativa entre classes mesmo com diferentes níveis de ajuste, atingindo níveis comparáveis à acurácia do diagnóstico clínico reportada na literatura, que situa-se tipicamente entre 80% e 90% (KLÖPPEL et al., 2008).

5.2.3.2 Divergência de Comportamento entre Datasets

A divergência de desempenho absoluto entre os resultados do ADNI e MIRIAD na Etapa 2 oferece insights valiosos sobre a interação entre estratégias de *fine-tuning* e características dos datasets:

Primeiro, o tamanho e complexidade dos dados. O ADNI, sendo um dataset multicêntrico maior (6400 amostras de 4 classes vs. 69 pacientes binários no MIRIAD) com maior heterogeneidade nas aquisições de imagem e na população de pacientes, beneficia-se substancialmente do *fine-tuning* completo, que permite aos modelos adaptarem suas representações internas à variabilidade específica do domínio médico. A ResNet152 com Full Fine-Tune (92.1%) demonstra que arquiteturas mais profundas conseguem melhor capturar esta complexidade quando todas as camadas são ajustadas.

Segundo, a relação entre tamanho do dataset e estratégia de *fine-tuning*. O MIRIAD, sendo menor e mais homogêneo (aquisições controladas de um único centro), apresenta resultados competitivos tanto com estratégias conservadoras (Top Layers: 83.2%) quanto com Full Fine-Tune (77.2%). A menor variabilidade intra-classe do MIRIAD permite que estratégias mais conservadoras sejam suficientes para adaptar os modelos ao domínio específico sem necessitar ajuste de todas as camadas.

Terceiro, a importância da combinação de técnicas. A comparação entre Etapa 1 (97.7% no ADNI multiclasse) e Etapa 2 Full Fine-Tune (99.0% no ADNI com EfficientNet) demonstra que diferentes estratégias de *fine-tuning*, quando aplicadas com o mesmo pipeline de otimização (data augmentation, early stopping, class weights e validação cruzada), podem alcançar e até superar resultados de abordagens mais simples. Isso valida que a escolha da estratégia de descongelamento de camadas é um fator adicional importante que interage com as demais técnicas de otimização para maximizar o desempenho.

5.2.3.3 Implicações Práticas e Recomendações

Os resultados da Etapa 2 fornecem diretrizes práticas importantes sobre a aplicação de *fine-tuning* em datasets médicos de neuroimagem:

Para datasets grandes e heterogêneos (tipo ADNI):

1. A estratégia Full Fine-Tune demonstrou ser a abordagem mais eficaz, alcançando 92.1% de acurácia com ResNet152. Esta estratégia permite que o modelo adapte todas

as suas representações internas ao domínio médico específico, superando amplamente estratégias conservadoras.

2. Estratégias parciais (Custom FC, Frozen, Top Layers) mostraram-se insuficientes para este contexto, sugerindo que datasets grandes e complexos requerem ajuste profundo para maximizar o desempenho.

3. A combinação de Full Fine-Tune com as técnicas de otimização da Etapa 1 (data augmentation, early stopping, validação cruzada) representa a configuração ideal, conforme evidenciado pelos 97.7% da Etapa 1.

Para datasets menores e homogêneos (tipo MIRIAD):

1. Estratégias conservadoras como Top Layers demonstraram eficácia notável (83.2%), oferecendo a melhor relação custo-benefício ao ajustar apenas as camadas finais mantendo as representações de baixo nível aprendidas no ImageNet.

2. Full Fine-Tune também apresentou resultados competitivos (77.2%), validando sua aplicabilidade mesmo em datasets menores, embora com ganhos marginais sobre estratégias mais conservadoras.

3. Para datasets desta escala, a escolha entre Top Layers e Full Fine-Tune pode ser guiada por considerações práticas como tempo de treinamento e recursos computacionais, dado que ambas apresentam desempenho clinicamente aceitável.

Diretrizes Gerais para Aplicações Clínicas:

1. Métricas balanceadas como prioridade: Os resultados evidenciam que acurácia isolada deve ser complementada com MCC, sensibilidade e especificidade para garantir desempenho robusto em ambas as classes. O MCC superior a 0.6 no MIRIAD e 0.87 no ADNI indica forte correlação entre predições e diagnósticos reais.

2. Validação rigorosa: A validação cruzada por paciente combinada com conjunto de teste independente (20% dos pacientes) demonstrou-se essencial para avaliar generalização real. A separação estrita por paciente evita vazamento de informação (data leakage).

3. Combinação de técnicas: Os melhores resultados (Etapa 1: 97.7% no ADNI) foram obtidos combinando *fine-tuning* completo com data augmentation, early stopping e validação cruzada, evidenciando que técnicas de otimização trabalham sinergicamente.

4. Transparência metodológica: Reportar resultados de diferentes estratégias, incluindo aquelas que não atingiram desempenho ótimo, contribui para o avanço do campo e evita vieses de publicação.

5.3 Comparação com a Literatura

A comparação direta com trabalhos da literatura apresenta desafios metodológicos importantes, uma vez que a maioria dos estudos no ADNI focam em tarefas de classificação binária (AD vs CN) ou triclasse (AD vs MCI vs CN), enquanto este trabalho aborda uma classificação multiclasse de 4 categorias diagnósticas que incluem gradações de severidade da demência.

Tabela 10 – Comparação de Desempenho com Trabalhos Relacionados

Estudo	Arquitetura	Tarefa	Métrica	Valor
Este Trabalho (Etapa 1)	EfficientNet-V2-M	4-class ADNI	Acurácia	0.977
El-Assy et al. (2024)	CNN Ensemble	4-class ADNI	Acurácia	0.996
Salehi et al. (2020)	CNN	Binária ADNI	Acurácia	0.990
Lu et al. (2022)	Inception-ResNet-V2	Binária ADNI	Acurácia	0.909
Feng et al. (2019)	3D-CNN + LSTM	Binária	Acurácia	0.948

Fonte: O autor (2025).

Este trabalho posiciona-se próximo ao estado da arte de (EL-ASSY et al., 2024), que também abordaram a classificação em quatro classes no ADNI, alcançando 99.6% de acurácia. A acurácia de 97.7% obtida neste estudo é notável considerando o severo desbalanceamento das classes (50% **EMCI**, 1% **CN**) e a dificuldade inerente de distinguir entre estágios adjacentes da doença.

Comparações com trabalhos de classificação binária (Salehi et al., Lu et al., Feng et al.) devem ser interpretadas com cautela, pois a tarefa multiclasse é intrinsecamente mais complexa, exigindo que o modelo aprenda fronteiras de decisão entre 4 categorias em vez de apenas 2.

O resultado do MIRIAD na Etapa 2 com ResNet152 e Top Layers (83.2% em tarefa binária) é comparável metodologicamente a (LU et al., 2022) que reportaram 90.9% no ADNI (também tarefa binária). Já este trabalho alcançou 97.7% no ADNI para tarefa multiclasse de 4 classes, significativamente mais desafiadora.

Ao utilizar Inception-ResNet-V2 com validação cruzada entre datasets esta comparação se torna particularmente relevante, pois (LU et al., 2022) estabeleceram um padrão rigoroso de generalização ao validar em datasets completamente distintos do treinamento. O desempenho obtido neste trabalho, validado com separação estrita por paciente e conjunto de teste independente, demonstra robustez metodológica comparável.

Em relação a (FENG et al., 2019), que alcançaram 94.8% utilizando uma arquitetura híbrida 3D-CNN + LSTM, é importante notar que sua abordagem explora informações temporais longitudinais (múltiplas aquisições do mesmo paciente ao longo do tempo), enquanto este trabalho utiliza apenas aquisições *cross-sectional* (imagens únicas por pa-

ciente). A arquitetura temporal oferece vantagens inerentes para capturar progressão da doença, mas exige dados longitudinais nem sempre disponíveis na prática clínica.

Os resultados aqui obtidos são validados por uma metodologia rigorosa que inclui: (i) validação cruzada estratificada por paciente em 5 folds, (ii) conjunto de teste final independente contendo 20% dos pacientes nunca vistos durante treinamento ou validação, e (iii) reporte transparente de todas as estratégias testadas, demonstrando que Full Fine-Tune alcançou 99.0% no ADNI na Etapa 2. Esta transparência metodológica confere alta confiança estatística aos achados e contribui para o avanço responsável do campo de diagnóstico auxiliado por IA em neuroimagem.

6 Conclusão

Este trabalho se propôs a desenvolver e avaliar sistematicamente um conjunto de modelos de *Deep Learning* para a classificação automatizada da Doença de Alzheimer a partir de imagens de ressonância magnética. Por meio de uma metodologia experimental estruturada em três estágios, buscou-se não apenas identificar as arquiteturas mais eficazes, mas também quantificar o impacto de diferentes técnicas de otimização e ajuste fino. Este capítulo final sintetiza os resultados obtidos, resgata os objetivos propostos e sugere direções para futuras pesquisas na área.

6.1 Síntese dos Resultados e Resgate dos Objetivos

O objetivo geral de comparar diferentes arquiteturas e estratégias de otimização foi plenamente alcançado. A análise revelou que, em um cenário de linha de base (Etapa 0) no dataset MIRIAD, uma CNN customizada apresentou o melhor desempenho. Contudo, ao introduzir técnicas de otimização avançadas como aumento de dados e validação cruzada (Etapa 1), as arquiteturas pré-treinadas EfficientNet-V2-M e ResNet152 emergiram como as mais robustas e com bom desempenho, especialmente no dataset ADNI, onde alcançaram um desempenho excepcional, com acurácias e F1-Scores médios superiores a 97%. Este ganho de desempenho entre os estágios evidencia a importância crucial de um pipeline de treinamento robusto para o sucesso de modelos de diagnóstico.

Um dos achados mais significativos e contra-intuitivos do trabalho surgiu na Etapa 2. O objetivo de explorar estratégias de *fine-tuning* foi cumprido, porém os resultados demonstraram que um ajuste fino mais profundo e granular foi prejudicial à capacidade de generalização dos modelos. Todas as estratégias testadas levaram a um sobreajuste severo, com um colapso no desempenho no conjunto de teste. Os resultados da Etapa 2 confirmaram que a estratégia Full Fine-Tune é altamente eficaz quando adequadamente implementada, alcançando 92.1% no ADNI e validando a hipótese de que modelos de *Deep Learning* com *fine-tuning* completo podem atingir alta precisão no diagnóstico da DA. A comparação entre as etapas evidencia que a combinação de *fine-tuning* completo com técnicas avançadas de otimização (Etapa 1) representa a configuração ideal para máximo desempenho.

Em síntese, o trabalho validou a hipótese de que modelos de *Deep Learning* podem atingir alta precisão no diagnóstico da DA. As arquiteturas EfficientNet-V2-M e ResNet152, quando devidamente otimizadas, se mostraram ferramentas de alto potencial. Os resultados, competitivos com o estado da arte, demonstram a viabilidade de desenvolver

sistemas automatizados de auxílio diagnóstico.

6.2 Propostas para Trabalhos Futuros

Apesar dos resultados promissores, este trabalho possui limitações que abrem caminho para diversas linhas de pesquisa futuras. As principais propostas de aprimoramento incluem:

Adoção de Protocolos 3D Completos: A principal limitação deste estudo foi o uso de fatias 2D. Futuros trabalhos devem focar na implementação de arquiteturas 3D (como 3D-CNNs ou Vision Transformers 3D) que possam processar o volume completo da ressonância magnética, preservando a rica informação espacial e volumétrica que é crucial para a análise da atrofia cerebral.

Aprimoramento da Interpretabilidade (XAI): Para superar a barreira da “caixa-preta” e aumentar a confiança clínica, é essencial integrar técnicas de Explainable AI. A aplicação de métodos como Grad-CAM, Saliency Maps ou LRP (Layer-wise Relevance Propagation) permitiria visualizar as regiões do cérebro que os modelos consideram mais importantes para a classificação, fornecendo insights valiosos para os médicos.

Integração Multimodal de Dados: A precisão diagnóstica pode ser significativamente aprimorada pela fusão de diferentes fontes de informação. Uma evolução natural deste trabalho seria a criação de modelos que integrem dados de MRI com outras modalidades, como imagens de PET, dados clínicos (testes neuropsicológicos) e biomarcadores genéticos.

Validações Multi-Centro e Longitudinais: Para testar verdadeiramente a robustez e a aplicabilidade clínica de um modelo, é necessário validá-lo em dados de múltiplos centros e hospitais, que possuam diferentes scanners e protocolos de aquisição. Além disso, a aplicação dos modelos em dados longitudinais, acompanhando pacientes ao longo do tempo, seria fundamental para avaliar sua capacidade de prever a progressão do MCI para a DA.

Avaliação de Custo-Efetividade Clínica: Por fim, um passo crucial para a translação da pesquisa para a prática clínica seria a realização de estudos de custo-efetividade, avaliando como a integração de uma ferramenta de diagnóstico automatizado impactaria o fluxo de trabalho clínico, o tempo para o diagnóstico e os custos para o sistema de saúde.

Referências

- ADNI Clinical Core. The alzheimer's disease neuroimaging initiative clinical core. *Alzheimer's & Dementia*, August 2024. > 2400 participants enrolled since 2004.
- ALRUILY, M. et al. Ensemble deep learning for alzheimer's disease diagnosis using mri scans. *PLOS ONE*, 2025. Ensemble of VGG16, MobileNet, InceptionResNetV2.
- Alzheimer's Association. 2024 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, v. 20, n. 5, p. 3708–3821, 2024.
- Alzheimer's Disease Neuroimaging Initiative (ADNI). *ADNI Database*. 2024. Disponível em: <<http://adni.loni.usc.edu/>>.
- CHAIRI, N. et al. Early alzheimer's diagnosis: Key to better care. *Clarivate Analytics Health Report*, July 2024. Discusses diagnostic challenges and stigma.
- DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2009. p. 248–255.
- DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. In: *International Conference on Learning Representations (ICLR)*. [S.l.: s.n.], 2021. Vision Transformer architecture.
- EL-ASSY, A. M. et al. A novel cnn architecture for accurate early detection and classification of alzheimer's disease using mri data. *Scientific Reports*, v. 14, n. 1, p. 3463, 2024.
- EL-GENEEDY, M. et al. An mri-based deep learning approach for accurate detection of alzheimer's disease. *Alexandria Engineering Journal*, v. 65, p. 89–101, 2023.
- FAROOQ, A. et al. A deep cnn based multi-class classification of alzheimer's disease using mri. In: *Proceedings of the IEEE International Conference on Imaging Systems and Techniques (IST)*. Beijing: IEEE, 2017. p. 1–6.
- FENG, C. et al. Deep learning framework for alzheimer's disease diagnosis via 3d-cnn and fsbi-lstm. *IEEE Access*, v. 7, p. 63605–63618, 2019.
- GOEDERT, M. et al. Neurofibrillary tangles and beta-amyloid deposits in alzheimer's disease. *Journal of Neural Transmission*, v. 33, p. 1–14, 1991. Citado por 175.
- GOENKA, N.; TIWARI, S. Deep learning for alzheimer prediction using brain biomarkers. *Artificial Intelligence Review*, v. 54, 10 2021.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge: MIT Press, 2016.
- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778. Citado por 293,000+.

- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.]: IEEE, 2016. p. 770–778. Winner of ILSVRC 2015, introduces residual connections.
- HEMMY, L. S. et al. Brief cognitive tests for distinguishing clinical alzheimer-type dementia from mild cognitive impairment or normal cognition in older adults. *Annals of Internal Medicine*, v. 172, n. 10, p. 678–687, 2020. Citado por 82.
- JACK, C. R. et al. NIA-AA research framework: Toward a biological definition of alzheimer’s disease. *Alzheimer’s & Dementia*, v. 14, n. 4, p. 535–562, 2018.
- JO, T.; NHO, K.; SAYKIN, A. J. Deep learning in alzheimer’s disease: Diagnostic classification and prognostic prediction using neuroimaging data. *Frontiers in Aging Neuroscience*, v. 11, p. 220, 2019. Citado por 784+.
- JUGANAVAR, A.; JOSHI, A. M. *Alzheimer’s Disease: Diagnosis and Challenges*. 2023. Current Medical Review. Discusses diagnostic tools limitations.
- KLÖPPEL, S. et al. Automatic classification of mr scans in alzheimer’s disease. *Brain*, Oxford University Press, v. 131, n. 3, p. 681–689, 2008.
- KRUTHIKA, K. R.; RAJESWARI, H. D.; MAHESHAPPA, H. D. Cbir system using capsule networks and 3d cnn for alzheimer’s disease diagnosis. *Informatics in Medicine Unlocked*, v. 14, p. 59–68, 2019.
- LANE, C. A.; HARDY, J.; SCHOTT, J. M. Alzheimer’s disease. *European Journal of Neurology*, v. 25, n. 1, p. 59–70, 2018.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Classic CNN architecture.
- LU, B. et al. A practical alzheimer’s disease classifier via brain imaging-based deep learning on 85,721 samples. *Journal of Big Data*, v. 9, n. 101, 2022.
- MALONE, I. B. et al. Miriad—public release of a multiple time point alzheimer’s mr imaging dataset. *NeuroImage*, v. 70, p. 33–36, 2013. 708 scans from 46 AD patients and 23 controls.
- MASTERS, C. L. et al. Alzheimer’s disease. *Nature Reviews Disease Primers*, v. 1, p. 15056, 2015.
- MISTUR, R. et al. Current challenges for the early detection of alzheimer’s disease: Brain imaging and csf studies. *Journal of Clinical Neurology*, v. 5, n. 4, p. 153–166, 2009. Citado por 134.
- NYU Langone Health. *Alzheimer’s Disease Neuroimaging Initiative*. 2025. NYU Data Catalog. Details on ADNI-1, ADNI-GO, ADNI-2, ADNI-3 phases. Disponível em: <<https://datacatalog.med.nyu.edu/dataset/10200>>.
- PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, v. 22, n. 10, p. 1345–1359, 2010. Classic transfer learning survey, cited by 30,000+.

- PATIL, V. et al. Patch-based classification for alzheimer disease using smri. In: *Proceedings of the IEEE International Conference*. [S.l.: s.n.], 2022. p. 1–6. 99.79% accuracy on MIRIAD dataset using 3D CNN.
- PEREZ, L. et al. Computer assisted diagnosis of alzheimer’s disease using statistical likelihood-ratio test. *PLOS ONE*, v. 18, n. 2, p. e0282731, February 2023. 89.1% sensibilidade e 87.0% especificidade no MIRIAD.
- PETERSEN, R. C. et al. Alzheimer’s disease neuroimaging initiative (adni): Clinical characterization. *Neurology*, v. 74, n. 3, p. 201–209, 2010. ADNI-1 baseline: 819 subjects (229 CN, 398 MCI, 192 AD).
- PETRELLA, J. R. et al. Mri-based deep learning predicts markers of alzheimer’s disease. *Radiology*, May 2024. RSNA News. Disponível em: <<https://www.rsna.org/news/2024/may/mri-deep-learning-alzheimers>>.
- RAGHU, M. et al. Transfusion: Unified medical image segmentation and classification with transformers. *arXiv preprint arXiv:1911.08272*, 2019.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. [S.l.: s.n.], 2015. p. 234–241. Citado por 70,000+.
- SALEHI, A. W. et al. A cnn model: Earlier diagnosis and classification of alzheimer disease using mri. In: *Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC)*. Trichy, India: IEEE, 2020. p. 156–161.
- SAOUD, L. S.; ALMARZOUQI, H. Explainable early detection of alzheimer’s disease using rois and an ensemble of 138 3d vision transformers. *Scientific Reports*, Nature Publishing Group, v. 14, n. 1, p. 27756, November 2024. Vision Transformer ensemble approach for Alzheimer’s disease detection. Disponível em: <<https://www.nature.com/articles/s41598-024-78861-5>>.
- SHAMRAT, F. M. J. M. et al. Alzheimernet: An effective deep learning based proposition for alzheimer’s disease stages classification from functional brain changes in magnetic resonance images. *IEEE Access*, v. 11, p. 16376–16395, 2023.
- SILVA, K. d. et al. Prediction of alzheimer’s disease from magnetic resonance imaging using deep learning. *Journal of Medical Imaging and Health Informatics*, 2023. Utilizou MIRIAD para validação, citado por 33+.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations (ICLR)*. [S.l.: s.n.], 2015. Citado por 100,000+.
- SOROUR, S. E. et al. Classification of alzheimer’s disease using mri data based on deep learning techniques. *Journal of King Saud University – Computer and Information Sciences*, 2024. Artigo 101940.
- TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning (ICML)*. [S.l.: s.n.], 2019. p. 6105–6114. Introduced compound scaling method.

TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*. PMLR, 2019. v. 97, p. 6105–6114. Introduces compound scaling method for CNNs. Disponível em: <<http://proceedings.mlr.press/v97/tan19a.html>>.

TUFAIL, A. B.; MA, Y.-K.; ZHANG, Q.-N. *Binary Classification of Alzheimer's Disease using sMRI imaging modality and Deep Learning*. 2018. ArXiv preprint. Disponível em: <<https://arxiv.org/pdf/1809.06209>>.

VINUKONDA, E. R. et al. An integrated deep learning model for early and multi-class alzheimer's disease classification using mri. *Scientific Reports*, v. 15, p. 10933, 2025. ADNI multi-class: 98.12% accuracy.

YOSINSKI, J. et al. How transferable are features in deep neural networks? In: *Advances in Neural Information Processing Systems 27 (NeurIPS)*. Curran Associates, Inc., 2014. p. 3320–3328. Seminal work on feature transferability in deep learning. Disponível em: <<https://arxiv.org/abs/1411.1792>>.

ZHUANG, F. et al. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, v. 109, n. 1, p. 43–76, 2021. Comprehensive modern survey.