

# Generalização de um classificador de Alzheimer em imagens de ressonância magnética para um banco de dados independente utilizando Transfer Learning

Bruno Porto Monteiro<sup>1</sup>, Gabriel Justino Panza<sup>1</sup>, Taiane Coelho Ramos<sup>1</sup>  
for the Alzheimer's Disease Neuroimaging Initiative<sup>2</sup>

<sup>1</sup> Instituto de Computação  
Universidade Federal Fluminense (UFF) – Niterói, RJ – Brazil

bpmonteiro@id.uff.br, gabrielpanza@id.uff.br, taianeramos@id.uff.br

**Resumo.** Modelos de Aprendizado Profundo (AP) têm sido propostos para auxílio ao diagnóstico de Alzheimer em imagens de Ressonância Magnética (MRI), alcançando altas taxas de acurácia quando avaliados no mesmo conjunto de dados. No entanto, quando aplicados a conjuntos externos, esses modelos não conseguem generalizar o aprendizado, tornando-se inviáveis na prática clínica. Esse problema é especialmente relevante quando hospitais substituem seus equipamentos de MRI por modelos mais novos, de outras marcas ou potências distintas. Nesse cenário, mesmo que já exista um modelo treinado no equipamento anterior, ele não pode ser reutilizado no novo, e o volume reduzido de dados coletados no equipamento novo é insuficiente para treinar um novo modelo do zero. Para lidar com esse desafio, este trabalho propõe o uso de Transfer Learning (TL) combinado com aumento de dados baseado na física de aquisição de MRI: a partir de um modelo com bom desempenho em um banco de dados extenso (ADNI), realizamos o ajuste fino para um banco de dados menor e independente (OASIS). O modelo adaptado por TL alcançou 89% de acurácia, representando um ganho de 18 pontos percentuais em relação ao modelo treinado diretamente no conjunto OASIS.

**Abstract.** Deep Learning (DL) models have been proposed to support Alzheimer's disease diagnosis in Magnetic Resonance Imaging (MRI), achieving high accuracy rates when evaluated on the same dataset. However, when applied to external datasets, these models often fail to generalize, hindering their clinical application. This issue is particularly relevant when hospitals upgrade MRI equipment to newer models, different brands, or distinct magnetic field strengths. In such scenarios, existing models trained on legacy equipment cannot be directly reused, and the limited volume of data collected from the new equipment is insufficient to train a model from scratch. To address this challenge, this study proposes the use of Transfer Learning (TL) combined with data augmentation based on MRI acquisition physics. Starting from a high-performance model trained on an extensive dataset (ADNI), we performed fine-tuning on a smaller, independent dataset (OASIS). The model adapted via TL achieved 89% accuracy, representing an 18-percentage-point gain compared to the model trained directly on the OASIS dataset.

## 1. Introdução

A doença de Alzheimer (AD), principal tipo de demência global com previsão de 79 milhões de casos até 2030 [International 2023], acarreta severos déficits cognitivos [Liu et al. 2021]. Terapias como *Lecanemab* e *Donanemab* são eficazes apenas em estágios iniciais, e a intervenção precoce pode retardar o declínio em 27% [Wang et al. 2024], tornando o diagnóstico precoce urgente. Neste contexto, Modelos de Aprendizado Profundo (AP) em imagens de Ressonância Magnética (MRI) surgem como solução, atingindo acurácias superiores a 95% [Turrisi et al. 2025].

Contudo, o desempenho desses modelos degrada drasticamente ao ser testado em dados de domínios diferentes (*domain shift*), inviabilizando a prática clínica quando há um modelo um modelo precisar ser usado em outras instituições ou quando há atualização de equipamentos [Turrisi et al. 2025, Yoon et al. 2024]. Como a generalização total exigiria milhares de dados de diversas instituições [Yoon et al. 2024], novos *scanners* com características técnicas distintas tornam modelos anteriores obsoletos [Litjens et al. 2017]. Assim, a adaptação a protocolos e hardwares variados permanece um desafio crítico na área.

Como contribuição, propomos um modelo de AP baseado em *Transfer Learning* (TL) para reaproveitar conhecimento prévio e generalizar para novos domínios via *fine-tuning* [Turrisi et al. 2025, Liu et al. 2021]. Para contornar a escassez de dados, comum em cenários clínicos reais, utilizamos aumento de dados (*data augmentation*) fundamentado na física de aquisição. Essa estratégia expõe o modelo a uma maior diversidade de condições, maximizando a extração de informação durante o ajuste fino.

## 2. Contribuições

Como contribuições deste trabalho, investigamos se TL combinado com aumento de dados melhora a generalização de modelos de AP para diagnóstico de Alzheimer em domínios externos. Especificamente, avaliamos se um modelo pré-treinado no ADNI pode ser adaptado ao OASIS, coletado por instituição diferente, com equipamentos e protocolos distintos e volume amostral reduzido. Nossa hipótese é que o TL, ao transferir conhecimento de grandes conjuntos de pesquisa, permite generalização para domínios independentes, mesmo com dados limitados, enquanto o aumento de dados atua como estratégia auxiliar ao simular variações técnicas de diferentes equipamentos.

As principais contribuições são: (i) desenvolvimento de pipeline completo de TL com pré-treinamento no ADNI e ajuste fino no OASIS, incluindo protocolo de congelamento de camadas e validação cruzada estratificada ( $k = 5$ ); (ii) avaliação sistemática do impacto de transformações geométricas e de intensidade em diferentes volumes de expansão amostral (1n, 4n, 8n, 12n), identificando o melhor compromisso entre desempenho e estabilidade; e (iii) validação externa em conjunto completamente independente, fornecendo evidências sobre capacidade real de generalização em cenários que simulam aplicação clínica em hospitais com recursos limitados.

## 3. Trabalhos Relacionados

Analisando a literatura recente, selecionamos estudos publicados nos últimos três anos que empregam arquiteturas de Redes Neurais Convolucionais (CNN) ou de *Transformers* para a detecção de AD e apresentaram propostas em bases de dados consolidadas,

como ADNI e OASIS, ou bancos da plataforma *Kaggle*, dos quais hajam informações disponíveis. As informações referentes a estes trabalhos estão disponíveis na Tabela 1.

A análise dos trabalhos da Tabela 1 reforça o argumento citado na seção anterior. Podemos visualizar métricas de acurácia superiores a 98% na maior parte dos trabalhos. Contudo, boa parte dos estudos avalia o desempenho exclusivamente na mesma base de treinamento, ainda que com divisões *train/test* distintas. Sendo assim, é desafiador avaliar a real capacidade de generalização dos modelos propostos, já que estes podem ter se adaptado excessivamente aos vieses do *dataset* treinado (*overfitting*).

**Tabela 1. Trabalhos relacionados sobre detecção de Alzheimer. No TL o modelo é pré-treinado em um banco de dados e passa por ajuste fino em um novo banco de dados. Nosso trabalho propõe o uso de TL para adaptar um modelo inicial para um banco de dados totalmente independente (OASIS).**

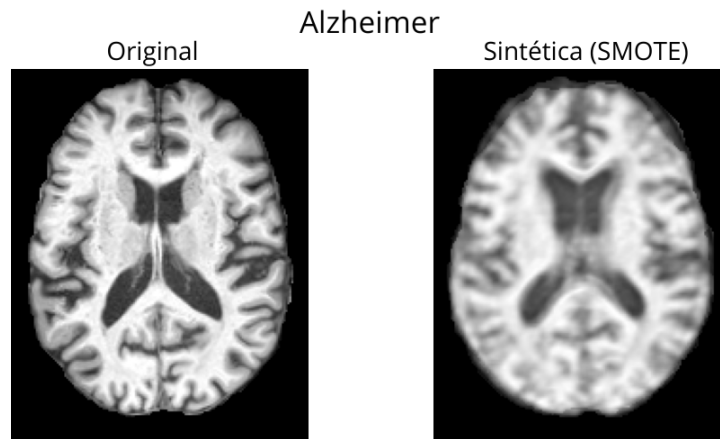
AUTORES	DATASET DE TREINO	VALIDAÇÃO INTERNA	AUMENTO DE DADOS	DATASET ALVO	ACURÁCIA COM TL
<i>Validação Interna — treino e teste no mesmo dataset</i>					
[Tang et al. 2023]	ADNI	94,22%	×	ADNI	×
[Salman et al. 2023]	OASIS	99,06%	Transformações	OASIS	×
[El-Assy et al. 2024]	ADNI	99,70%	Dados Artificiais	ADNI	×
[Khan et al. 2024]	ADNI	98,00%	×	ADNI	×
[Al-Dhief et al. 2024]	Kaggle	99,00%	×	Kaggle	×
[Hassan et al. 2025]	ADNI/OASIS/Kaggle	99,00%	Dados Artificiais	ADNI/OASIS/Kaggle	×
[Mousavi et al. 2025]	Kaggle	96,89%	Transformações	Kaggle	×
[Oraby et al. 2025]	Kaggle	99,22%	Dados Artificiais	Kaggle	×
[Dardouri 2025]	Kaggle	99,68%	Transformações	Kaggle	×
[Sriram et al. 2025]	Kaggle	94,27%	×	Kaggle	×
[Fang et al. 2025]	ADNI	98,60%	×	ADNI	×
<i>Transfer Learning — adaptação de modelo para dataset externo</i>					
[Turrisi et al. 2025]	ADNI (1.5T)	99,00%*	Transformações	ADNI (3T)	98,75%*
<b>Este trabalho (2026)</b>	<b>ADNI</b>	<b>96,00%</b>	<b>Transformações</b>	<b>OASIS</b>	<b>88,96%</b>

\*Resultados não reprodutíveis: ao seguir fielmente o código-fonte e os dados disponibilizados pelos autores, obtivemos acurácia de 0,63 e *F1-Score* de 0,59 ao aplicar o TL.

Outro ponto identificado na literatura foi que diversos estudos utilizam a geração de dados artificiais, baseados em algoritmos como o SMOTE, para mitigar o desbalançamento de classes. Contudo, estes métodos podem gerar dados artificiais distantes dos padrões reais de MRI [Perez-Garcia et al. 2021], com perdas anatômicas e de qualidade. Na Figura 1, ilustramos um exemplo dessa degradação, pois a amostra gerada por SMOTE apresenta qualidade abaixo, além de perda de definição na região do córtex pré-frontal.

Turrisi et al. (2025) investigaram o uso de TL para mitigar problemas de troca de domínio decorrentes da evolução tecnológica de equipamentos de MRI, adaptando modelos entre diferentes intensidades de campo magnético (1,5T e 3T) dentro da base de dados ADNI, reportando acurácia de 99%. Embora represente um avanço em relação aos trabalhos anteriores, a adaptação permanece intra-domínio no mesmo dado, o que dificulta avaliar a capacidade de generalização da proposta.

Portanto, apresentamos uma contribuição científica para a comunidade ao avaliar um modelo pré-treinado no ADNI e verificar sua generalização para um domínio independente, utilizando o conjunto de dados OASIS. Ao avaliar um domínio completamente independente, propomos uma abordagem mais próxima da prática clínica. Para isso, empregamos transformações que mimetizam artefatos de diferentes *scanners*, priorizando a robustez diante da variabilidade técnica real, em detrimento de acurácias em bases únicas.



**Figura 1.** Exemplo de amostra real (esquerda) e amostra sintética gerada por SMOTE para a classe de Alzheimer (direita). A amostra sintética apresenta artefatos e menor qualidade, evidenciando que a interpolação entre voxels não preserva a coerência de aquisição de um MRI.

## 4. Métodos

Este trabalho investiga a capacidade de generalização de modelos de AP entre domínios distintos por meio de TL. Um modelo é treinado no conjunto ADNI ( $n = 1493$ ) e adaptado ao conjunto externo OASIS ( $n = 163$ ) por meio de *fine-tuning*. Também avaliamos o impacto de diferentes estratégias de aumento de dados na robustez diante da troca de domínio. A seguir, apresentamos a metodologia completa deste trabalho.

### 4.1. Conjuntos de Dados

#### 4.1.1. Alzheimer's Disease Neuroimaging Initiative (ADNI)

Os dados utilizados no pré-treinamento foram coletados pela *Alzheimer's Disease Neuroimaging Initiative* (ADNI) [Jack et al. 2008]. Foram selecionadas imagens de *MRI-3D T1*, adquiridas com intensidade de campo de *1,5 Tesla*. Para garantir a qualidade do sinal, inspecionamos visualmente, de forma criteriosa, para descartar volumes que apresentavam artefatos de movimento severos ou borrões. Ao fim do processo, o conjunto final foi composto por 1493 indivíduos (727 mulheres e 766 homens), com idades entre 54 e 93 anos (média  $74,7 \pm 8,73$ ). As amostras são provenientes das bases *ADNI GO*, *ADNI1* e *ADNI2*, categorizadas em saudável (CN), com *Clinical Dementia Rating* (CDR) de 0, e Alzheimer (AD), com CDR de 1.

#### 4.1.2. Open Access Series of Imaging Studies (OASIS)

Para a avaliação da adaptação de domínio, usamos a primeira versão do conjunto *Open Access Series of Imaging Studies* (OASIS). Mais especificamente, utilizamos os dados do OASIS-1 [Marcus et al. 2007]. Este conjunto, composto por imagens *MRI 3D T1* de *1,5 Tesla* já pré-processadas, de forma independente do ADNI, por especialistas associados ao projeto OASIS, apresenta desafios inerentes à troca de domínio devido a protocolos de aquisição distintos. Os dados são classificados pelo índice, em que CDR 0.0

indica ausência de demência (cognitivamente normal, CN), da qual temos 135 amostras, e  $CDR \geq 1.0$  indica demência compatível com Alzheimer (AD), com 28 amostras.

## 4.2. Pré-processamento

Para o conjunto ADNI, desenvolvemos um protocolo próprio de pré-processamento composto por: registro estereotático ao Atlas *ICBM 152 Nonlinear atlases version 2009* [Fonov et al. 2011] por transformações afins, extração de tecido cerebral (*skull stripping*), recorte volumétrico para eliminar *voxels* nulos, correção de viés de campo, *winsorization* (percentis 1 e 99) e normalização *min-max* no intervalo  $[0, 1]$ . Para evitar contaminação entre conjuntos, utilizamos uma MRI por indivíduo, com distribuição estratificada em 70% treino (CN: 525, AD: 530), 20% validação (CN: 137, AD: 152) e 10% teste (CN: 73, AD: 76).

O conjunto OASIS foi utilizado já pré-processado pelos responsáveis do projeto, por meio de um *pipeline* baseado no FSL (*FMRIB Software Library*), que inclui extração de tecido cerebral e normalização de intensidade. Aplicamos apenas a normalização *min-max* para o intervalo  $[0, 1]$ . Devido ao desbalanceamento amostral e ao tamanho reduzido ( $n = 163$ ), adotamos validação cruzada estratificada ( $k = 5$ ), garantindo a proporção original de classes em cada partição. Para mitigar o desbalanceamento, aplicamos *oversampling* exclusivamente no conjunto de treino de cada partição, replicando amostras da classe AD até igualar a classe CN, evitando vazamento de dados.

## 4.3. Arquitetura CNN-3D

Neste trabalho, propomos uma CNN-3D com 3 blocos convolucionais (4, 8 e 16 filtros de tamanho  $3 \times 3 \times 3$ ). A arquitetura foi definida por busca empírica, e as estratégias de regularização e ativação foram selecionadas com base em suas propriedades conhecidas para dados volumétricos com volume amostral reduzido. Cada bloco é seguido por *Batch Normalization*, *Leaky ReLU* ( $\alpha = 0,3$ ), *Average Pooling* e *Dropout* de 0,3. Após o *Flatten*, o modelo possui uma camada densa de 16 neurônios com regularização *L2* (0,01), *Batch Normalization*, *Dropout* de 0,3 e *Leaky ReLU*, antes da saída *softmax*. No pré-treinamento (ADNI), utilizamos *Adam* com taxa de aprendizado de 0,0001, até 200 épocas e *Early Stopping* (paciência de 20 épocas). No ajuste fino (OASIS), a taxa de aprendizado foi de 0,0005, até 100 épocas e *Early Stopping* (paciência de 15 épocas).

## 4.4. Experimentos

Para investigar de forma sistemática o impacto da troca de domínio e do TL na classificação de doença de Alzheimer em exames de *MRI*, estruturamos um conjunto de experimentos controlados. Cada experimento foi realizado para isolar um fator específico do problema, permitindo avaliar separadamente: (1) a capacidade de reprodução de um método existente da literatura, (2) o desempenho do modelo quando treinado em um conjunto de dados grande e homogêneo, (3) as limitações de treinamento em bases pequenas, (4) o grau de generalização direta entre bases distintas e (5) o impacto de uma estratégia de adaptação baseada em TL combinada com aumento de dados. Dessa forma, os experimentos, ilustrados na Figura 2, foram organizados de modo progressivo, partindo de cenários controlados dentro de uma mesma base até chegar à proposta otimizada para lidar com troca de domínio entre bases distintas.

1. **Reprodução de Turrisi et al. (2025):** Reproduzimos o trabalho de [Turrisi et al. 2025], que investiga TL para adaptação entre aquisições de *MRI* a *1,5T* e *3T* dentro do ADNI, utilizando o código-fonte e os indivíduos disponibilizados pelos autores no GitHub<sup>1</sup>. A reprodução consistiu no ajuste fino da rede *ADnet* — pré-treinada a *1,5T* ( $n = 550$ ) e adaptada ao domínio alvo a *3T* ( $n = 80$ ) — com congelamento das duas primeiras camadas e aumento de dados por transformações afins (rotação, *zoom* e translação), gerando seis imagens por exame no conjunto de treino de cada *fold* ( $k = 5$ ).
2. **ADNI → ADNI (Baseline):** Treinamento e avaliação realizados exclusivamente no ADNI, sem TL, nem aumento de dados, estabelecendo uma referência de desempenho intra-domínio e permitindo verificar a consistência dos resultados com a literatura.
3. **OASIS → OASIS (Treinamento Local Isolado):** Treinamento e avaliação realizados exclusivamente no OASIS — um conjunto menor —, sem uso de TL, mas com balanceamento de classes. Este experimento simula o treinamento de um modelo por uma clínica, em que há poucos dados anotados disponíveis.
4. **ADNI → OASIS sem TL (Generalização Direta):** Modelo treinado com ADNI, mas avaliado diretamente no banco OASIS, sem aplicação de TL, nem aumento de dados, permitindo mensurar o impacto isolado da troca de domínio sem qualquer adaptação.
5. **ADNI → OASIS com TL (Proposta):** Experimento central do trabalho. O modelo pré-treinado no ADNI é adaptado ao OASIS por ajuste fino, com avaliação sistemática de estratégias de aumento de dados em diferentes fatores de expansão, descritos na Seção 4.6.

Nos experimentos apresentados, diferentes estratégias de avaliação foram empregadas de acordo com o tamanho das bases de dados. Os experimentos cuja base de treinamento principal é o ADNI (itens 2 e 4) adotaram a estratégia de *holdout*, aproveitando o maior volume de dados disponível. Por outro lado, os experimentos que envolvem treinamento no OASIS (itens 3 e 5) utilizaram validação cruzada estratificada ( $k = 5$ ), devido ao número reduzido de amostras. No item 1, a validação cruzada também foi mantida para garantir a reprodução fiel da metodologia descrita por [Turrisi et al. 2025].

#### 4.5. Transfer Learning

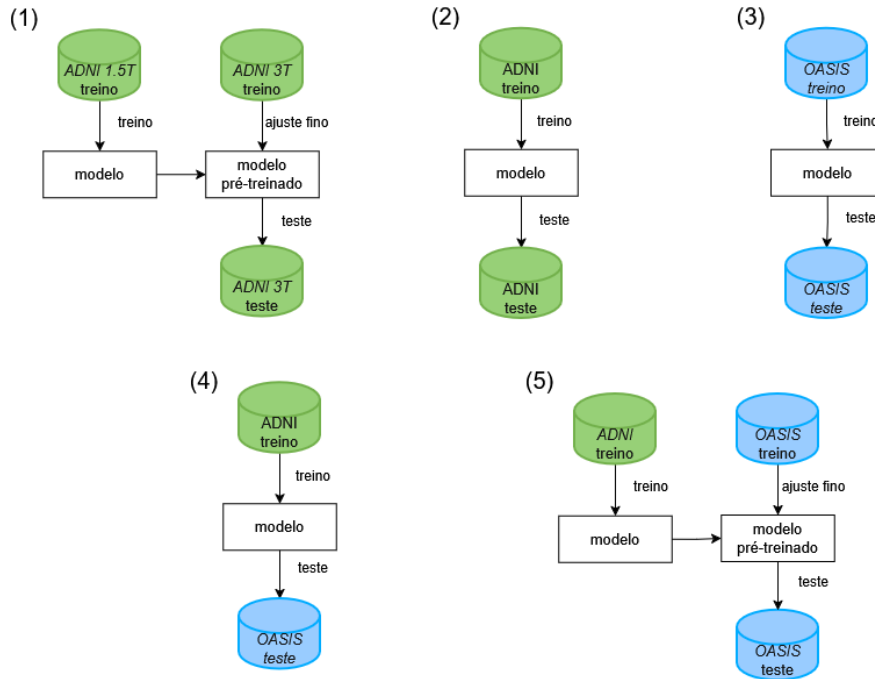
O protocolo de TL consistiu no congelamento do primeiro bloco convolucional completo — camada de entrada, convolução, *Batch Normalization*, *Leaky ReLU*, *Average Pooling* e *Dropout*, totalizando 6 camadas. Camadas iniciais de CNNs tendem a aprender representações de baixo nível transferíveis entre domínios [Yosinski et al. 2014], de forma que seu congelamento preserva o conhecimento morfológico aprendido no ADNI durante o ajuste fino em um conjunto menor. As camadas restantes foram reajustadas com taxa de aprendizado de 0,0005 com decaimento adaptativo, e uma nova camada de saída foi adicionada para classificação binária no OASIS.

#### 4.6. Aumento de Dados

O aumento de dados *on-the-fly* foi aplicado via *TorchIO* [Perez-Garcia et al. 2021], tanto no conjunto de treino do pré-treinamento (ADNI)

---

<sup>1</sup><https://github.com/rturrisige/TL4ADdiagnosis/>



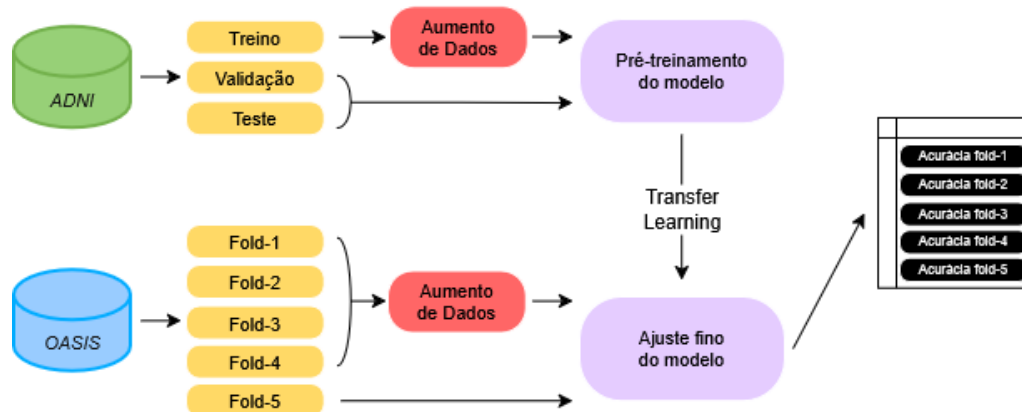
**Figura 2. Esquema dos cinco experimentos realizados. (1) Reprodução do trabalho de Turrisi et al (2025). (2) Baseline: treinamento e avaliação no ADNI. (3) Treinamento local isolado: treinamento e avaliação no OASIS. (4) Generalização direta: treinamento no ADNI e avaliação no OASIS sem adaptação. (5) Proposta principal (esquema geral): pré-treinamento no ADNI e ajuste fino no OASIS, repetido para cada configuração de aumento de dados (1n, 4n, 8n e 12n, com transformações geométricas e de intensidade).**

quanto no ajuste fino (OASIS), exclusivamente no conjunto de treino de cada *fold*. As transformações foram organizadas em dois grupos: (i) apenas geométricas (rotação, translação e *zoom*); e (ii) geométricas + intensidade (ruído gaussiano, campo de viés e artefatos de movimento), reproduzindo artefatos técnicos de diferentes equipamentos.

Para avaliar o impacto dos dados sintéticos, cruzaram-se dois escopos de transformação com fatores de expansão de  $1n$  (substituição das 130 amostras originais),  $4n$ ,  $8n$  e  $12n$  (preservação das originais com acréscimo de versões sintéticas, totalizando até 1560 imagens). Em todos os cenários, a sobreamostragem da classe AD foi realizada dentro de cada *fold* após a divisão dos dados, mitigando o risco de contaminação do conjunto de teste.

#### 4.7. Pipeline proposto

O pipeline proposto, ilustrado na Figura 3, é composto por duas etapas sequenciais. Na primeira, o modelo é pré-treinado no conjunto ADNI, com divisão estratificada em treino, validação e teste, e aumento de dados aplicado exclusivamente no conjunto de treino. O melhor modelo obtido nessa etapa é então transferido para a segunda etapa, na qual é adaptado ao conjunto OASIS por meio de ajuste fino com validação cruzada estratificada ( $k = 5$ ). Em cada partição, o aumento de dados é aplicado exclusivamente no conjunto de treino do respectivo *fold*, garantindo que o modelo seja exposto à variabilidade técnica de diferentes equipamentos sem contaminação do conjunto de teste.



**Figura 3. Pipeline proposto: o modelo é pré-treinado no ADNI com separação treino/validação/teste e aumento de dados. O melhor modelo é então adaptado ao OASIS via ajuste fino com validação cruzada estratificada ( $k = 5$ ), onde o aumento de dados é aplicado exclusivamente nos folds de treino.**

#### 4.8. Métricas de Avaliação

Duas métricas principais foram adotadas para avaliar o desempenho dos experimentos. A acurácia agregada consolida as predições de todos os  *folds*  em um único conjunto, calculando a proporção global de acertos e refletindo o desempenho médio do modelo ao longo das partições. A  *Macro Average F1-Score*  calcula a média não ponderada do  *F1-Score*  entre as classes, atribuindo peso igual a cada uma independentemente do desbalanceamento amostral.

### 5. Resultados

#### 5.1. Experimentos de Referência

O experimento (1), reprodução do trabalho de Turrise et al. (2025), resultou em acurácia global de 0,63 e  *F1-Score*  de 0,71 para a classe AD, desempenho substancialmente inferior aos 99% reportados pelos autores. O modelo apresentou baixo  *recall*  para a classe CN (40,63%), classificando erroneamente a maioria dos indivíduos saudáveis como portadores de AD.

O experimento (2), ADNI  $\rightarrow$  ADNI, atingiu acurácia de 0,96 e  *F1-Score*  de 0,96, confirmando que o modelo apresenta bom desempenho para classificação da doença quando treinado e avaliado no mesmo domínio.

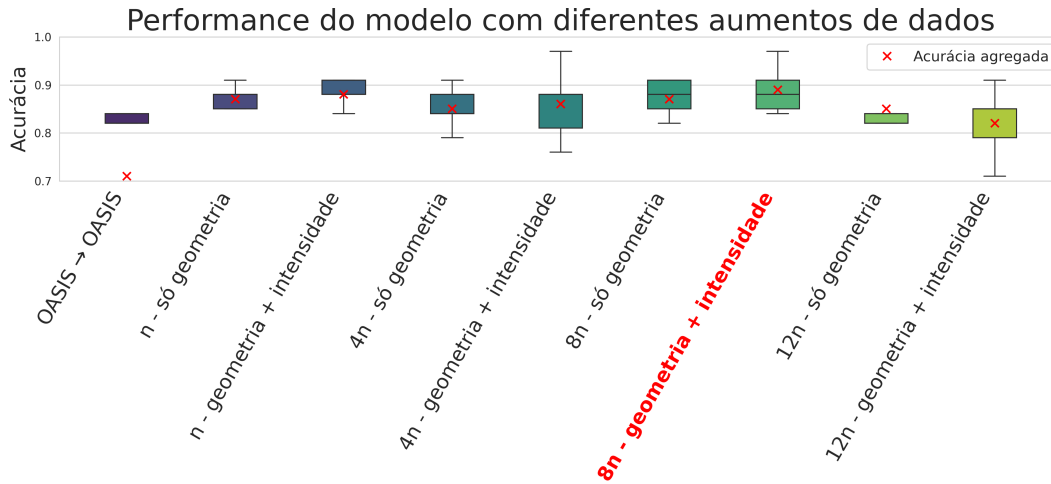
O experimento (3), OASIS  $\rightarrow$  OASIS sem TL, alcançou acurácia de 0,71 e  *F1-Score*  de 0,23 para a classe AD, com sensibilidade de apenas 0,25, evidenciando a insuficiência do treinamento isolado com volume reduzido de dados.

O experimento (4), ADNI  $\rightarrow$  OASIS sem TL, em que o modelo treinado no ADNI foi aplicado diretamente ao OASIS sem adaptação, alcançou acurácia de 0,80 e  *F1-Score*  de 0,58, indicando que o conhecimento adquirido no ADNI generaliza parcialmente para o domínio externo, mas de forma insuficiente para uso clínico.

#### 5.2. TL com Ajuste Fino

A Figura 4 apresenta a acurácia obtida em cada  *fold*  do experimento (5) para cada configuração de aumento de dados, incluindo o experimento (3), OASIS  $\rightarrow$  OASIS, como

referência de modelo sem TL.



**Figura 4.** Boxplots de acurácia para experimento (5), que utiliza TL com estratégias de aumento de dados e validação cruzada ( $k = 5$ ). Cada boxplot é gerado a partir do valor de acurácia no conjunto teste de cada *fold* e a cruz vermelha indica a acurácia agregada de todos os *fold*s. O primeiro boxplot à esquerda representa o experimento (3) OASIS → OASIS e foi incluído para comparação com um experimento sem TL. Este possui um ponto de valor 0,15, resultando em uma queda da acurácia média deste experimento. Destacamos em vermelho o experimento 8n (Geometria + Intensidade), que obteve a melhor acurácia agregada entre *fold*s, com valor de 0,89.

No fator 1n, a configuração Geometria alcançou acurácia de 0,87 e *F1-Score* de 0,51, enquanto a configuração Geometria + Intensidade obteve acurácia de 0,88 e *F1-Score* de 0,56, demonstrando que a inclusão de transformações de intensidade contribuiu para a detecção da classe AD já no menor fator de expansão.

No fator 4n, a configuração Geometria alcançou acurácia de 0,85 e *F1-Score* de 0,52, e a configuração Geometria + Intensidade obteve acurácia de 0,86 e *F1-Score* de 0,55. O aumento no fator de expansão não resultou em ganho expressivo em relação ao 1n, sugerindo que o volume amostral isolado não oferece melhoras expressivas.

No fator 8n, a configuração Geometria alcançou acurácia de 0,87 e *F1-Score* de 0,60. A configuração Geometria + Intensidade obteve o melhor desempenho geral, com acurácia de 0,89, *Macro Avg* de 0,81 e *F1-Score* de 0,68, com sensibilidade de 0,68 para a classe AD. Essa configuração representa o melhor equilíbrio entre volume amostral e diversidade de transformações.

No fator 12n, observou-se degradação do desempenho. A configuração Geometria caiu para acurácia de 0,71 e *F1-Score* de 0,52, e a configuração Geometria + Intensidade alcançou acurácia de 0,84 e *F1-Score* de 0,62. O excesso de variações sintéticas introduziu ruído excessivo, sugerindo um limite de saturação para o aumento de dados neste conjunto.

**Tabela 2. Métricas quantitativas dos experimentos. Todas as métricas referem-se à classe AD. O experimento 8n (Geometria + Intensidade) apresenta melhor equilíbrio entre todas as métricas. As métricas Precisão, Recall e *F1-Score* são referentes à classe Alzheimer.**

Experimento	Acurácia Agregada	Macro Avg	Precisão	Recall	<i>F1-Score</i>
<i>Experimentos de Referência (1 - 4)</i>					
1. Reprodução de Turrisi et al (2025)	0,63	0,59	0,66	0,77	0,71
2. ADNI → ADNI	0,96	0,96	0,99	0,97	0,96
3. OASIS → OASIS (sem TL)	0,71	0,52	0,21	0,25	0,23
4. ADNI → OASIS (sem TL)	0,80	0,73	0,46	0,79	0,58
<i>Experimento 5 - TL com Ajuste Fino</i>					
1n (Geometria)	0,87	0,72	0,73	0,39	0,51
1n (Geometria + Intensidade)	0,88	0,75	0,80	0,43	0,56
4n (Geometria)	0,85	0,72	0,59	0,46	0,52
4n (Geometria + Intensidade)	0,86	0,73	0,62	0,50	0,55
8n (Geometria)	0,87	0,76	0,64	0,57	0,60
<b>8n (Geometria + Intensidade)</b>	<b>0,89</b>	<b>0,81</b>	<b>0,68</b>	<b>0,68</b>	<b>0,68</b>
12n (Geometria)	0,71	0,66	0,36	0,89	0,52
12n (Geometria + Intensidade)	0,84	0,76	0,54	0,71	0,62

## 6. Discussão

### 6.1. TL como Estratégia de Adaptação de Domínio

Os resultados confirmam que o TL é o fator determinante para a generalização no domínio externo: enquanto o treinamento isolado no OASIS e a aplicação direta do modelo ADNI sem adaptação resultaram em desempenho insuficiente para uso clínico, o ajuste fino ao domínio-alvo produziu ganhos consistentes em todas as configurações avaliadas. O aumento de dados mostrou-se uma estratégia auxiliar eficaz, com ganhos progressivos até o fator 8n (Geometria + Intensidade), a partir do qual o excesso de variações sintéticas introduziu ruído excessivo, sugerindo um limite de saturação. Contudo, a variabilidade entre partições (0,82–0,97) ressalta que validação em conjuntos independentes de maior porte é essencial antes de considerar aplicação clínica.

Em comparação com a literatura, os trabalhos da Tabela 1 reportam acurácias superiores a 98%, porém avaliados exclusivamente no mesmo domínio de treinamento, o que não permite aferir a real capacidade de generalização. O trabalho de Turrisi et al. (2025), único a propor TL na literatura analisada, reporta acurácia de 98,75% em cenário intra-domínio dentro do próprio ADNI. Contudo, ao reproduzirmos o experimento fielmente, obtivemos acurácia de apenas 0,63 e *F1-Score* de 0,59, sugerindo que os resultados originais podem não ser reproduzíveis. Nosso trabalho opera em um cenário de troca para um domínio totalmente independente, tornando os 89% alcançados um resultado mais robusto e clinicamente relevante do que acurácias nominais em bases únicas.

### 6.2. Limitações e Trabalhos Futuros

Neste trabalho, apenas um banco de dados externo foi utilizado para avaliar a eficácia do pipeline frente à troca de domínio. Além disso, o pipeline não incorpora técnicas de explicabilidade, não oferecendo *insights* sobre as regiões cerebrais que fundamentam as predições. Como trabalhos futuros, propõe-se a avaliação em outros conjuntos externos para ampliar a variabilidade de domínios testados, além da aplicação de *Grad-CAM* para maior compreensão do comportamento do modelo.

## 7. Conclusão

Este trabalho demonstra que o *Transfer Learning* (TL) é o fator determinante para a generalização em domínios externos, permitindo que o conhecimento morfológico prévio compense a escassez de dados locais. A hipótese foi confirmada: a adaptação via ajuste fino, suportada por aumentos de dados baseados na física de aquisição, superou significativamente o treinamento isolado. Identificou-se que a eficácia desta adaptação é otimizada na configuração 8n (Geometria + Intensidade), que oferece robustez técnica sem atingir o limite de saturação e ruído observado no fator 12n. Conclui-se que o TL, validado em conjuntos independentes, é o caminho para viabilizar modelos de diagnóstico em ambientes clínicos heterogêneos.

## Agradecimentos

Este trabalho foi parcialmente financiado pela Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) E-26/210.759/2024 (295437).

## Referências

- Al-Dhief, F., Al-Hgarat, N., Faheem, M., Al-Shammaa, A., Al-Mqdadi, F., Sallam, A., Al-Saba, M. F., Al-Sammaraie, N., Natij, K., Qazi, A., Al-Maatouk, Q., and Khan, S. (2024). Leveraging transformers and explainable ai for alzheimer’s disease interpretability. *Scientific Reports*, 14(1):28864.
- Dardouri, S. (2025). An efficient method for early alzheimer’s disease detection based on mri images using deep convolutional neural networks. *Frontiers in Artificial Intelligence*, 8:1563016.
- El-Assy, A. M., Amer, H. M., Ibrahim, H. M., and Mohamed, M. A. (2024). A novel cnn architecture for accurate early detection and classification of alzheimer’s disease using mri data. *Scientific Reports*, 14:3463.
- Fang, W., Wei, Z., Yue, Y., Zhang, L., and Wu, S. (2025). An artificial intelligence-based framework for Alzheimer’s disease diagnosis from magnetic resonance imaging volumes via video vision transformer. *arXiv preprint arXiv:2501.15733*.
- Fonov, V. S., Evans, A. C., Botteron, K., Almlí, C. R., McKinstry, R. C., Collins, D. L., and Giedd, J. R. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1):1–11.
- Hassan, N., Miah, A. S. M., Suzuki, K., Okuyama, Y., and Shin, J. (2025). Stacked cnn-based multichannel attention networks for alzheimer disease detection. *Scientific Reports*, 15(1):5815.
- International, A. D. (2023). World alzheimer report. Acessado em: 14 de outubro de 2024.
- Jack, C. R., Bernstein, M. A., Fox, et al. (2008). The alzheimer’s disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691.
- Khan, A. A., Mahendran, R. K., Perumal, K., and Faheem, M. (2024). Dual-3dm3-ad: Mixed transformer based semantic segmentation and triplet pre-processing for early

- multi-class alzheimer's diagnosis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:696–707.
- Litjens, G., Kooi, T., Ehteshami Bejnordi, B., Adiyoso Setio, A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88.
- Liu, Q., Wang, L., Guo, Z., et al. (2021). Alzheimer's disease: Mechanisms and therapeutic strategies. *Frontiers in Neuroscience*, 15:714177.
- Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2007). Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498–1507.
- Mousavi, S. M., Moulaei, K., and Ahmadian, L. (2025). Classifying and diagnosing alzheimer's disease with deep learning using 6735 brain mri images. *Scientific Reports*, 15(1):22721.
- Oraby, S., Emran, A., El-Saghir, B., and Mohsen, S. (2025). Hybrid of dsr-gan and cnn for alzheimer disease detection based on mri images. *Scientific Reports*, 15(1):12727.
- Perez-Garcia, F., Sparks, R., and Ourselin, S. (2021). Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 208:106236.
- Salman, S. M., Abdul-Ghafour, A. H., Al-Sultani, H. K., Al-Ghazali, B. S. A., Al-Sultany, B. A. S., Al-Hussainy, A. K., Al-Sultani, K. D., Shah, S. U. N., S. S., M. B. H., Al-Daami, A. H. M., Al-Sanjary, O. I., Al-Anee, A. F. M., and Al-Lami, A. S. H. (2023). Artificial cognition for detection of mental disability: A vision transformer approach for alzheimer's disease. *Applied Sciences*, 13(16):9322.
- Sriram, S., Nivethitha, V., Arun Kaarthic, T. P., Archita, S., and Murugan, T. (2025). Advanced mri based alzheimer's diagnosis through ensemble learning techniques. *Scientific Reports*, 15(1):33840.
- Tang, C., Wei, M., Chen, J., Zhang, Y., and Luo, Y. (2023). CsAGP: Detecting Alzheimer's disease from multimodal images via dual-transformer with cross-attention and graph pooling. *Journal of King Saud University-Computer and Information Sciences*, 35(7):101618.
- Turrise, R., Pati, S., Pioggia, G., and Tartarisco, G. (2025). Adapting to evolving mri data: A transfer learning approach for alzheimer's disease prediction. *NeuroImage*, 307:121016.
- Wang, Y., Gao, R., Wei, T., Johnston, L., Yuan, X., Zhang, Y., and Yu, Z. (2024). Predicting long-term progression of alzheimer's disease using a multimodal deep learning model incorporating interaction effects. *Journal of Translational Medicine*, 22:265.
- Yoon, J. S., Oh, K., Shin, Y., Mazurowski, M. A., and Suk, H.-I. (2024). Domain generalization for medical image analysis: A review. *arXiv preprint arXiv:2310.08598*.
- Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, volume 27.