Revista Eletrônica
Acervo Saúde
Electronic Journal Collection Health ISSN 2178-2091

Classificação de estágios da doença de Alzheimer em imagens de ressonância magnética por redes neurais convolucionais

Classification of Alzheimer's disease stages in magnetic resonance imaging using convolutional neural networks

Clasificación de las etapas de la enfermedad de Alzheimer en imágenes de resonancia magnética mediante redes neuronales convolucionales

Rafael Farias Batista[1], Bianka dos Santos Gouvêa[1], Lizandra de Lima Quaresma[1], Marta de Oliveira Barreiros[1].

RESUMO

Objetivo: Propor um algoritmo inteligente para diagnosticar a Doença de Alzheimer usando exames de ressonância magnética, além de comparar o desempenho de cinco modelos de redes neurais convolucionais (ResNet50, VGG19, DenseNet201, InceptionV3 e EfficientNetB7) na classificação automática de estágios da doença. **Métodos:** Nesta pesquisa, foi feio um estudo experimental realizado com 416 exames de ressonância magnética funcional oriundos do dataset OASIS-1, envolvendo pacientes entre 18 e 96 anos. Foram utilizadas imagens segmentadas entre as fatias 100 e 160, com pré-processamento padronizado e treinamento dos modelos utilizando o otimizador Adam (taxa de aprendizado 10^{-3}, batch size 32), com parada antecipada e redução adaptativa da taxa de aprendizado. **Resultados:** Na classificação multiclasses (Sem Demência, Demência ?Muito Leve, Demência Leve e Demência Moderada), ResNet50 obteve acurácia de 83%, DenseNet201 de 82% e VGG19 de 80%, destacando-se frente ao InceptionV3 (78%) e EfficientNetB7 (72%). A abordagem hierárquica binária (com demência e sem demência) demonstrou ganhos em precisão, com destaque para a ResNet50, que atingiu acurácia de 96%. **Conclusão:** As redes convolucionais apresentam desempenho promissor para o diagnóstico assistido da Doença de Alzheimer, especialmente com a adoção de estratégias hierárquicas que reduzem erros na distinção entre os estágios da doença.

Palavras-chave: Doença de Alzheimer, Redes neurais convolucionais, Ressonância magnética.

ABSTRACT

Objective: To propose an intelligent algorithm for diagnosing Alzheimer's Disease using magnetic resonance imaging (MRI), as well as to compare the performance of five convolutional neural network models (ResNet50, VGG19, DenseNet201, InceptionV3, and EfficientNetB7) in the automatic classification of the disease stages. **Methods:** In this research, an experimental study was conducted with 416 functional magnetic resonance imaging exams from the OASIS-1 dataset, involving patients aged between 18 and 96 years. Segmented images between slices 100 and 160 were used, with standardized preprocessing and model training using the Adam optimizer (learning rate 10^{-3}, batch size 32), with early stopping and adaptive learning rate reduction. **Results:** In the multiclass classification (No Dementia, Very Mild Dementia, Mild Dementia, and Moderate Dementia), ResNet50 achieved an accuracy of 83%, DenseNet201 82%, and VGG19 80%, standing out compared to InceptionV3 (78%) and EfficientNetB7 (72%). The binary hierarchical approach (with dementia

[1] Universidade do Estado do Pará (UEPA), Castanhal - PA.

SUBMETIDO EM: 6/2025 | **ACEITO EM: 7/2025** | **PUBLICADO EM: 8/2025**

and without dementia) showed gains in accuracy, especially ResNet50, which achieved an accuracy of 96%. **Conclusion:** Convolutional networks show promising performance for assisted diagnosis of Alzheimer's Disease, especially with the adoption of hierarchical strategies that reduce errors in distinguishing the stages of the disease.

Keywords: Alzheimer's disease, Convolutional neural networks, Magnetic resonance imaging.

RESUMEN

Objetivo: Proponer un algoritmo inteligente para diagnosticar la Enfermedad de Alzheimer mediante resonancia magnética, y comparar el desempeño de cinco modelos de redes neuronales convolucionales (ResNet50, VGG19, DenseNet201, InceptionV3 y EfficientNetB7) en la clasificación automática de sus etapas. **Métodos:** Se realizó un estudio experimental con 416 exámenes de resonancia magnética funcional del conjunto de datos OASIS-1, con pacientes entre 18 y 96 años. Se utilizaron imágenes segmentadas entre los cortes 100 y 160, con preprocesamiento estandarizado y entrenamiento de los modelos usando el optimizador Adam (tasa de aprendizaje 10^{-3}, batch size 32), con parada anticipada y reducción adaptativa del aprendizaje. **Resultados:** En la clasificación multiclase (Sin Demencia, Demencia Muy Leve, Leve y Moderada), ResNet50 alcanzó una precisión del 83%, DenseNet201 del 82% y VGG19 del 80%, superando a InceptionV3 (78%) y EfficientNetB7 (72%). El enfoque jerárquico binario (con y sin demencia) mostró mejoras, destacando ResNet50 con 96% de precisión. **Conclusión:** Las redes convolucionales muestran un desempeño prometedor para el diagnóstico asistido del Alzheimer, especialmente con estrategias jerárquicas que reducen errores al distinguir las etapas de la enfermedad.

Palabras clave: Enfermedad de Alzheimer, Redes neuronales convolucionales, Resonancia magnética.

INTRODUÇÃO

A Doença de Alzheimer (DA) é uma enfermidade neurodegenerativa progressiva que afeta predominantemente a população idosa, sendo responsável por cerca de 70% dos casos de demência (CASTELLANI RJ, et al., 2010). Estima-se que atualmente mais de 35 milhões de pessoas no mundo vivem com DA, o que impõe um elevado ônus social e econômico para sistemas de saúde e famílias (GUZIOR N, et al., 2015). Embora o diagnóstico clínico baseado em avaliações neuropsicológicas seja amplamente utilizado, técnicas de imagem, em especial a Ressonância Magnética (RM), têm se mostrado fundamentais para a detecção precoce das alterações estruturais cerebrais características da doença (CHANDRA A, et al., 2019; JACK CR JR, et al., 1997). No entanto, a interpretação manual desses exames é demorada e sujeita a variabilidade interobservadora, limitando sua aplicabilidade em larga escala.

As redes Neurais Convolucionais (*Convolutional Neural Network* - CNNs) despontam como ferramentas promissoras para a análise automatizada de imagens médicas, permitindo extrair padrões sutis relacionados à atrofia cortical e ao hipocampo mesmo em estágios iniciais da DA (ESTEVA A, et al., 2017; LITJENS G, et al., 2017). Apesar do avanço em classificações binárias (paciente versus não-paciente), poucos estudos abordam a categorização multiclasses dos diferentes níveis de comprometimento cognitivo, o que é crucial para orientar intervenções terapêuticas e prognósticos mais precisos (EBRAHIMI A, et al., 2021). Dessa forma, as CNNs têm se destacado na área médica por sua capacidade de processar imagens e auxiliar em diagnósticos clínicos. Com isso, estudos mostram sua eficácia na detecção de doenças, como câncer, pneumonia e doenças cardiovasculares, por meio da análise de exames de imagem (LITJENS et al., 2017).

Recentemente, as técnicas de aprendizado profundo têm sigo foram empregadas na classificação de neuroimagens cerebrais, sendo utilizado conjuntos de dados de ressonância magnética, como o ADNI, para distinguir entre os grupos: Comprometimento Cognitivo Leve (CCL), Doença de Alzheimer (DA), Cognitivamente Normal (CN) e Indivíduos Saudáveis (ARCHANA & KALIRAJAN, 2023). Em outro estudo, foi realizada a classificação e predição dos estágios da DA utilizando o conjunto de dados de ressonância magnética ADNI, por meio de uma rede neural convolucional. A pesquisa promoveu uma análise abrangente

dos dados de neuroimagem disponíveis, visando à identificação da DA (MISHRA et al., 2024). Em outro estudo, foram utilizadas três diferentes arquiteturas de redes convolucionais pré-treinadas (AlexNet, GoogleNet e MobileNetV2), comparando distintos otimizadores empregados na minimização da função de perda durante o treinamento dos modelos de aprendizado de máquina. Com isso, os resultados demonstraram que a aprendizagem por transferência é uma abordagem eficaz para superar as limitações dos modelos convencionais, que restringem a precisão do diagnóstico da DA (HUSSAIN, et al., 2025). Diante desses estudos, o uso de CNNs na medicina promete transformar o diagnóstico, e essa abordagem também pode ser aplicada na identificação de padrões relacionados à Doença de Alzheimer, permitindo uma detecção mais precoce e precisa da condição.

Nesse cenário, o presente trabalho tem como objetivo avaliar o desempenho de cinco arquiteturas de CNN amplamente utilizadas na literatura — ResNet50, VGG-19, DenseNet201, InceptionV3 e EfficientNetB7 — na classificação dos estágios da Doença de Alzheimer a partir de imagens de RM provenientes do banco de dados OASIS-1. Ao comparar o desempenho dessas redes em um mesmo contexto experimental, busca-se identificar lacunas na capacidade de discriminação entre os estágios sem demência, demência muito leve, leve e moderado da DA, contribuindo para o aprimoramento de sistemas de auxílio ao diagnóstico precoce e para o avanço do conhecimento sobre aplicações de aprendizado profundo em neuroimagem.

MÉTODOS

Este estudo adota uma abordagem quantitativa com método experimental, onde as CNNs são aplicadas para classificar os estágios da DA com base em imagens de Ressonância Magnética (RM). O processo metodológico foi estruturado em etapas sequenciais, que serão descritas a seguir.

Neste trabalho foi usado o conjunto de dados públicos do repositório OASIS-1 (https://sites.wustl.edu/oasisbrains/home/oasis-1/), o qual disponibiliza uma coleção transversal de imagens de ressonância magnética estrutural de 416 indivíduos, com idades variando entre 18 e 96 anos, incluindo sujeitos com diagnóstico de DA em estágio inicial. Para cada participante foram adquiridas entre três e quatro varreduras individuais de ressonância magnética. Todos os indivíduos incluídos são destros, havendo representação de ambos os sexos.

Em termos de distribuição, o conjunto contempla 100 sujeitos com idade superior a 60 anos diagnosticados com DA em estágio muito leve a moderado. Adicionalmente, para 20 indivíduos sem demência foram incluídas imagens correspondentes a sessões de seguimento realizadas após um intervalo de 90 dias, com o objetivo de monitorar a estabilidade do quadro clínico. Importante destacar que o OASIS-1 também incorpora dados clínicos relevantes para a caracterização do estado cognitivo dos participantes.

Entre esses, destacam-se as escalas Mini-Mental State Examination (MMSE), conforme descrito por Rubin et al. (1998), e Clinical Dementia Rating (CDR), proposta por Morris (1993). A escala CDR classifica os indivíduos em diferentes estágios de demência: 0 = sem demência; 0.5 = demência muito leve; 1 = demência leve; 2 = demência moderada. Todos os participantes com CDR > 0 foram classificados com provável DA, conforme os critérios clínicos adotados no estudo original. Essas medidas fornecem uma referência clínica adicional para a validação dos modelos preditivos.

Além das imagens e dados clínicos, o conjunto de dados inclui informações demográficas, tais como sexo (masculino, feminino), idade, nível educacional e status socioeconômico. No entanto, tais variáveis não foram consideradas nas análises preditivas realizadas neste estudo, que busca explorar exclusivamente o potencial das redes neurais convolucionais para a detecção automática da DA com base em características extraídas das imagens. Como discutido por Peng L, et al. (2020), embora as informações demográficas possam fornecer dados valiosos em outras abordagens, este estudo foca nas características das imagens para promover uma análise mais centrada na patologia da DA.

Tabela 1 - Resumo dos dados demográficos dos indivíduos e do estado de demência. (M – Masculino, F – Feminino).

		Sem demência				Com demência				
Idade	**N**	**n**	**Média**	**M**	**F**	**n**	**Média**	**M**	**F**	**CDR 0.5/1/2**
<20	19	19	18.53	10	9	0	-	-	-	
20s	119	119	22.82	51	68	0	-	-	-	
30s	16	16	33.38	11	5	0	-	-	-	
40s	31	31	45.58	10	21	0	-	-	-	
50s	33	33	54.36	11	22	0	-	-	-	
60s	40	25	64.88	7	18	15	66.13	6	9	12/3/0
70	83	35	73.37	10	25	48	74.42	20	28	32/15/1
80s	62	30	84.07	8	22	32	82.88	13	19	22/9/1
>=90	13	8	91.00	1	7	5	92.00	2	3	4/1/0
Total	416	316		119	197	100		41	59	70/8/2

Fonte: Batista RF, et al., 2025. Dados extraídos de Kaggle - OASIS Alzheimer 's Detection.

Para acessibilidade, os arquivos originais (.img e .hdr) foram convertidos para o formato Nifti (.nii) com o FSL, abrangendo imagens de 416 pacientes. No treinamento da rede neural, foram utilizadas imagens 2D, obtidas pela segmentação do cérebro ao longo do eixo z em 256 fatias, das quais foram selecionadas as fatias de 100 a 160 para análise.

A escolha dos cortes 100-160 no dataset OASIS-1 é motivada pela sua relevância para capturar o hipocampo e regiões temporais mediais, que são biomarcadores estruturais cruciais no diagnóstico do Alzheimer devido à atrofia característica nessas áreas. Segundo Jack et al. (1997), a atrofia hipocampal é um dos primeiros. O dataset OASIS-1, descrito por Marcus et al. (2007), contém imagens T1-weighted com resolução de 1 mm³, e as slices coronais na faixa de 100 a 160 frequentemente abrangem o hipocampo e porções do lobo temporal medial, que são críticas para análise volumétrica e detecção de alterações patológicas. Estudos como o de LaMontagne et al. (2021) reforçam que essas regiões, visíveis em cortes coronais, são prioritárias para análises de neuroimagem em Alzheimer, justificando a seleção dessas fatias para maximizar a relevância dos dados processados

Os experimentos foram realizados no ambiente de desenvolvimento em nuvem oferecido pelo Google Colaboratory (Colab), que disponibiliza recursos computacionais gratuitos com suporte a unidades de processamento gráfico (GPU). Neste estudo, foi utilizada uma instância equipada com uma GPU NVIDIA Tesla T4, contendo 16 GB de memória dedicada, além de até 12 GB de RAM e cerca de 70 GB de armazenamento temporário, características que permitiram a execução eficiente de tarefas intensivas como o treinamento de redes neurais profundas.

A plataforma foi executada com Python 3.10 e as seguintes bibliotecas: NumPy 1.24.3 e Pandas 1.5.3 para manipulação de dados; OpenCV 4.7.0 e Pillow 9.5.0 para leitura e processamento de imagens; e Matplotlib 3.7.1 e Seaborn 0.12.2 para geração de visualizações gráficas. O desenvolvimento dos modelos de deep learning foi conduzido com a biblioteca TensorFlow 2.12.0, utilizando a API Keras 2.12.0 (ABADI, et al., 2016). Para pré-processamento, avaliação de desempenho e validação dos modelos, foi utilizada a biblioteca Scikit-learn 1.2.2 (VAN DER MOLEN, 2020). A utilização da GPU gratuita acelerou significativamente os processos de treinamento dos modelos, especialmente para arquiteturas convolucionais mais complexas como ResNet50, InceptionV3 e EfficientNetB7, amplamente empregadas em tarefas de classificação de imagens médicas (PASCANU, et al., 2013). A combinação entre a infraestrutura escalável do Google Colab e o conjunto atualizado de bibliotecas possibilitou a condução dos experimentos com alta reprodutibilidade, desempenho computacional eficiente e facilidade de uso, tornando o ambiente adequado para estudos automatizados baseados em imagens de ressonância magnética.

Os modelos de redes neurais convolucionais (CNNs) ResNet50, VGG19, DenseNet201, InceptionV3 e EfficientNetB7 representam avanços significativos na área de visão computacional, especialmente em tarefas de classificação de imagens. A VGG19, proposta por Simonyan e Zisserman (2014), destaca-se por sua arquitetura simples, composta por camadas convolucionais com filtros 3×3 e operações de pooling, o que

facilita sua implementação, embora resulte em elevado custo computacional. Em contraste, a ResNet50, desenvolvida por He et al. (2016), introduz conexões residuais que atenuam o problema do desaparecimento do gradiente, permitindo o treinamento eficiente de redes mais profundas. A DenseNet201, apresentada por Huang et al. (2017), contribui com uma estrutura inovadora ao conectar cada camada a todas as anteriores, promovendo maior reutilização de características e economia de parâmetros. A InceptionV3, de Szegedy et al. (2016), utiliza módulos com convoluções paralelas de diferentes tamanhos de kernel, o que favorece a extração multiescalar de padrões e melhora a eficiência computacional. Já a EfficientNetB7, desenvolvida por Tan e Le (2019), alcançou desempenho superior ao empregar um escalonamento composto, ajustando de forma equilibrada profundidade, largura e resolução. Esses modelos têm sido amplamente aplicados em diagnósticos médicos por imagem, como na detecção precoce de doenças neurológicas, devido à alta acurácia e capacidade de generalização.

Nos experimentos realizados, todos os modelos foram treinados com hiperparâmetros otimizados, utilizando o otimizador ADAM, taxa de aprendizado inicial de 10^{-3} e tamanho de lote de 32, visando maximizar o desempenho e reduzir o risco de overfitting. O treinamento foi configurado para ocorrer em até 50 épocas, com a aplicação da estratégia de parada antecipada (early stopping) após 10 épocas sem melhoria na perda de validação, conforme a condição, conforme a Equação 1:

$$L\_\{val\}\^\{(t)\} \geq L\_\{val\}\^\{(t-10)\} \quad (1)$$

onde L_{val}^{(t)} representa a perda de validação na época t. Adicionalmente, uma redução dinâmica da taxa de aprendizado (η) foi aplicada automaticamente quando a perda de validação não apresentava melhoria por cinco épocas consecutivas, sendo definida por:

$$\eta = \eta_0 \times \gamma^k \quad (2)$$

onde γ = 0,1 é o fator de redução e k é o número de vezes em que a redução foi aplicada. Esse ajuste contínuo dos hiperparâmetros foi realizado por meio de tentativa e erro, com monitoramento constante da perda de validação. A combinação dessas estratégias garantiu uma melhor generalização dos modelos, evitando o sobreajuste e assegurando a adaptação dos parâmetros de forma eficiente durante o treinamento, em conformidade com as práticas recomendadas para redes neurais convolucionais profundas (GOODFELLOW et al., 2016).

O modelo para classificação dos estágios do Alzheimer por imagens de ressonância magnética (RM) foi desenvolvido em três camadas principais: Dados, Treinamento e Testes. Na Camada de Dados, as imagens foram obtidas de um banco específico (como OASIS), aplicando-se técnicas de aumento de dados (rotação, inversão, escalonamento) para diversificar o conjunto e evitar o overfitting. Após o processamento, os dados foram divididos em treinamento, validação e teste.

Na Camada de Treinamento, foram utilizados modelos pré-treinados (ResNet50, VGG19, DenseNet201, InceptionV3 e EfficientNetB7), ajustados com otimizadores como ADAM. Além disso, para o desenvolvimento dos modelos de classificação, foi utilizado um conjunto de dados contendo imagens, igualmente distribuídas, com 475 imagens por classe, sendo que 76% ficaram para treinamento da rede, 19% para validação e 5% para teste. Ademais, foram usados critérios de aprendizado, usando o conjunto de validação, como ajuste dinâmico da taxa de aprendizado e parada antecipada, garantiram um treinamento eficiente. Como citado por Zhu et al. (2019), o uso de aprendizado profundo em imagens médicas continua a se expandir, oferecendo grandes avanços na precisão da detecção e no diagnóstico precoce, incluindo em condições como Alzheimer.

Tabela 2 - Quantidade de imagens do conjunto do OASIS-1 para treinamento, validação e teste dos modelos convolucionais.

Classificadores	Com demência	Sem demência			Total
		Muito Leve	Leve	Moderada	
Classificador Multiclasse	475	475	475	475	**1900**
Classificador Binário	475	475	475	475	**1900**
Multiclasse de demência	0	475	475	475	**1425**
	Total				**5225**

Fonte: Batista RF, et al., 2025.

Na **Tabela 3** está apresentado os hiperparâmetros utilizados pelo modelo. Os valores definidos para os hiperparâmetros utilizados nos experimentos com os modelos ResNet50, VGG19, DenseNet201, InceptionV3 e EfficientNetB7, bem como a metodologia aplicada para sua determinação, são apresentados a seguir. Essa padronização garante que os modelos sejam avaliados em um contexto consistente, permitindo uma comparação justa de seus desempenhos. Todas as execuções foram realizadas com o objetivo de maximizar a acurácia na base de validação, minimizando o risco de overfitting.

Tabela 3 - Hiperparâmetros utilizados por modelo. (TA – Taxa de aprendizado, TL – Tamanho do Lote, PA – Parada Antecipada, RTA – Redução da taxa de aprendizado)

Otimizador	TA	TL	Épocas	PA	RTA
ADAM	10^-3	32	50	10 Épocas	5 Épocas

Fonte: Batista RF, et al., 2025.

Por fim, na Camada de Teste, o modelo realizou a classificação das imagens entre as categorias, e, para isso, foi separado um conjunto independente de teste, sendo 25 imagens por classe, que não participaram do treinamento nem da validação. Este conjunto foi reservado exclusivamente para a avaliação final dos modelos após o término do processo de aprendizado. A utilização de um conjunto de teste independente é essencial para estimar a capacidade de generalização do modelo em dados inéditos, fornecendo uma métrica mais realista e imparcial do desempenho do sistema em cenários práticos. O desempenho foi avaliado por meio das métricas de acurácia, precisão, recall e F1-score, conforme amplamente recomendado na literatura (Goodfellow et al., 2016; Liu et al., 2020; Zhang et al., 2018). A acurácia representa a proporção de previsões corretas em relação ao total de previsões realizadas; a precisão indica a proporção de predições positivas corretas em relação ao total de predições positivas feitas pelo modelo; a recall mensura a capacidade do modelo em identificar corretamente os casos positivos, ou seja, a taxa de detecção dos verdadeiros positivos; e o F1-score corresponde à média harmônica entre precisão e recall, proporcionando um equilíbrio entre essas duas métricas.

As fórmulas dessas métricas são apresentadas pelas Equações a seguir:

$$\text{Acurácia} = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$

$$\text{Precisão} = TP / (TP + FP) \quad (4)$$

$$\text{Recall} = TP / (TP + FN) \quad (5)$$

$$\text{F1 score} = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (6)$$

RESULTADOS E DISCUSSÃO

Os resultados de avaliação dos modelos de dados estão apresentados na **Tabela 4**. De forma geral, os modelos DenseNet201 e VGG19 destacaram-se como as arquiteturas mais robustas, com acurácias de 82% e 80%, respectivamente. O ResNet50 apresentou acurácia ligeiramente superior (83%), no entanto, sua performance foi mais polarizada entre as classes, com desempenho elevado para casos sem demência e para demência muito leve, mas inferior para as outras classes. Já os modelos InceptionV3 e EfficientNetB7 apresentaram acurácias globais menores (78% e 72%, respectivamente), refletindo uma maior dificuldade na distinção dos diferentes estágios da doença (**Tabela 3**).

Tabela 4 - Resumo dos resultados da previsão sobre os dados de teste com os modelos usados no trabalho usando 4 classes de classificação.

Modelo	Classe	Precisão	Recall	F1-Score	Acurácia
ResNet50	Sem demência	0.93	0.93	0.93	**0.83**
	Demência muito leve	1.00	1.00	1.00	
	Demência leve	0.71	0.80	0.75	
	Demência moderada	0.69	0.60	0.64	
VGG19	Sem demência	0.73	0.73	0.73	**0.80**
	Demência muito leve	0.94	1.00	0.97	
	Demência leve	0.75	0.80	0.77	
	Demência moderada	0.77	0.67	0.71	
DenseNet201	Sem demência	0.82	0.93	0.88	**0.82**
	Demência muito leve	1.00	1.00	1.00	
	Demência leve	0.83	0.67	0.74	
	Demência moderada	0.62	0.67	0.65	
InceptionV3	Sem demência	0.67	0.93	0.78	**0.78**
	Demência muito leve	1.00	1.00	1.00	
	Demência leve	0.75	0.60	0.67	
	Demência moderada	0.75	0.60	0.67	
EfficientNetB7	Sem demência	0.61	0.73	0.67	**0.72**
	Demência muito leve	0.83	1.00	0.91	
	Demência leve	0.71	0.67	0.69	
	Demência moderada	0.70	0.47	0.56	

Fonte: Batista RF, et al., 2025.

Especificamente para a classe Sem demência, o ResNet50 apresentou o melhor resultado (F1-score = 0.93), seguido pela DenseNet201 (F1-score = 0.88). Isso indica uma alta capacidade desses modelos em identificar corretamente indivíduos saudáveis, com baixos índices de falsos positivos. Em contrapartida, o EfficientNetB7 apresentou o pior desempenho nessa classe (F1-score = 0.67), sugerindo uma maior taxa de erros nessa categoria. Já a classe Demência muito leve foi aquela em que todos os modelos apresentaram excelente desempenho. As arquiteturas ResNet50, VGG19, DenseNet201 e InceptionV3 alcançaram F1-score igual a 1.00, evidenciando uma elevada sensibilidade para detectar os primeiros sinais de comprometimento cognitivo. Apenas o EfficientNetB7 obteve uma leve redução nesta métrica (F1-score = 0.91), ainda assim mantendo um resultado satisfatório. Por outro lado, a classificação da Demência leve revelou-se mais desafiadora. A DenseNet201 foi o modelo que melhor lidou com esta categoria (F1-score = 0.74), seguida pela VGG19 (F1-score = 0.77). Os demais modelos, incluindo o EfficientNetB7 (F1-score = 0.69), demonstraram dificuldades em distinguir esta fase intermediária, o que pode ser atribuído à similaridade morfológica das imagens desta classe com as de outros estágios da doença. Por fim, a Demência moderada foi a classe com os piores resultados globais. O VGG19 obteve o melhor F1-score (0.71), seguido por InceptionV3 (0.67), enquanto o EfficientNetB7 apresentou o desempenho mais baixo (F1-score = 0.56). Esses resultados indicam uma redução significativa na sensibilidade dos modelos para detectar casos mais avançados da doença, possivelmente devido à menor representatividade dessa classe no conjunto de dados e à maior variabilidade das manifestações anatômicas.

Em contraponto à abordagem anterior de classificação de 4 classes, propôs-se também uma metodologia alternativa, fundamentada em uma estrutura hierárquica de classificação, porém usando a mesma divisão de dados. Nessa perspectiva, o processo classificatório foi dividido em duas etapas. Primeiramente, os dados foram segmentados de forma binária, separando os casos entre "sem demência" e "com demência" (englobando os estágios muito leve, leve e moderada). Em seguida, os casos classificados como "com demência" foram processados por um segundo modelo, responsável por realizar a classificação interna entre os subestágios: demência muito leve, demência leve e demência moderada. Essa abordagem visa reduzir a complexidade do processo classificatório, minimizar erros de confusão entre classes clinicamente próximas e refletir de maneira mais fiel o raciocínio diagnóstico utilizado na prática clínica, onde, primeiramente, é avaliada a presença ou ausência de demência, para então ser determinada a sua gravidade.

A **Tabela 5** apresenta o desempenho de cinco modelos CNN na classificação de pacientes com e sem demência. Todos os modelos mostraram alta acurácia, entre 87% e 92%, sendo o EfficientNet o melhor, com 92% de acurácia. Os modelos apresentaram maior facilidade em identificar pacientes sem demência, com precisão e recall geralmente acima de 90%. Para a classe “Demência”, o desempenho foi inferior, com recall variando entre 0.73 (VGG19) e 0.80 (outros modelos), indicando dificuldades na detecção completa dos casos positivos. O EfficientNet destacou-se por equilibrar melhor precisão e recall na classe demência, resultando em um F1-score superior. Esses resultados sugerem que, embora as redes consigam classificar bem, ainda há espaço para melhorar a sensibilidade na detecção da demência, essencial para diagnósticos confiáveis.

Tabela 5 - Resumo dos resultados da previsão sobre os dados de teste com os modelos usados no trabalho usando classificação binária.

Modelo	Classe	Precisão	Recall	F1-Score	Acurácia
ResNet50	Sem demência	0.90	0.93	0.92	**0.89**
	Demência	0.86	0.80	0.83	
VGG19	Sem demência	0.88	1.00	0.94	**0.91**
	Demência	1.00	0.73	0.85	
DenseNet201	Sem demência	0.93	0.93	0.93	**0.90**
	Demência	0.80	0.80	0.80	
InceptionV3	Sem demência	0.90	0.90	0.90	**0.87**
	Demência	0.80	0.80	0.80	
EfficientNetB7	Sem demência	0.93	0.96	0.95	**0.92**
	Demência	0.86	0.80	0.83	

Fonte: Batista RF, et al., 2025.

Em síntese, os resultados obtidos com a aplicação do modelo binário indicam que esta abordagem além de ser viável, é também eficaz como primeiro filtro no processo de diagnóstico automatizado, demonstrando potencial para ser integrada em sistemas de apoio à decisão clínica.

A **Tabela 6** apresenta a avaliação comparativa de cinco modelos. As acurácias globais variaram entre 0.81 (EfficientNet) e 0.96 (ResNet50), indicando desempenho variável entre as arquiteturas. O ResNet50 obteve a melhor performance geral, com acurácia de 96%, destacando-se por apresentar elevados índices de precisão e recall em todas as classes, especialmente na classe “Demência Muito Leve” (Precisão = 0.93; Recall = 1.00) e “Demência Leve” (Precisão = 0.96; Recall = 1.00). Apenas na “Demência Moderada” houve uma ligeira redução no recall (0.88), embora mantendo elevada precisão (1.00). O VGG19 e DenseNet apresentaram desempenhos semelhantes, com acurácias de 0.92 e 0.93, respectivamente. Ambas as arquiteturas classificaram perfeitamente a classe “Demência Leve” (Precisão e Recall = 1.00), mas mostraram redução na performance para a classe “Demência Moderada”, especialmente no recall (0.80). A classe “Demência Muito Leve” também apresentou variação: VGG19 com F1-Score de 0.89 e DenseNet com 0.93.O Inception obteve acurácia global de 0.91, com desempenho consistente na “Demência Leve” (Precisão = 0.96; Recall = 1.00), mas com menor efetividade na detecção das classes “Demência Muito Leve” (F1-Score = 0.88) e “Demência Moderada” (F1-Score = 0.85). Por outro lado, o EfficientNet apresentou o desempenho mais modesto, com acurácia de 0.81. Embora tenha mantido boa performance na classificação da “Demência Leve” (F1-Score = 0.96), apresentou dificuldades nas classes “Demência Muito Leve” e “Demência Moderada”, com F1-Score de 0.73 em ambas, decorrente de baixos valores de precisão e recall (~0.75 e ~0.72, respectivamente).

Tabela 6 - Resumo dos resultados da previsão sobre os dados de teste com os modelos usados no trabalho usando as classes de demência.

Modelo	Classe	Precisão	Recall	F1-Score	Acurácia
ResNet50	Demência Muito Leve	0.93	1.00	0.96	**0.96**
	Demência Leve	0.96	1.00	0.98	
	Demência Moderada	1.00	0.88	0.94	
VGG19	Demência Muito Leve	0.83	0.96	0.89	**0.92**
	Demência Leve	1.00	1.00	1.00	
	Demência Moderada	0.95	0.80	0.87	
DenseNet201	Demência Muito Leve	0.86	1.00	0.93	**0.93**
	Demência Leve	0.96	1.00	0.98	
	Demência Moderada	1.00	0.80	0.89	
InceptionV3	Demência Muito Leve	0.85	0.92	0.88	**0.91**
	Demência Leve	0.96	1.00	0.98	
	Demência Moderada	0.91	0.80	0.85	
EfficientNetB7	Demência Muito Leve	0.75	0.72	0.73	**0.81**
	Demência Leve	0.93	1.00	0.96	
	Demência Moderada	0.75	0.72	0.73	

Fonte: Batista RF, et al., 2025.

CONCLUSÃO

Os resultados demonstram que modelos mais profundos e complexos, como a ResNet50 e a DenseNet201, possuem maior capacidade discriminativa, especialmente quando combinados com uma abordagem hierárquica de classificação, que se mostrou vantajosa frente ao modelo multiclasses tradicional. A elevada precisão na detecção de estágios iniciais de demência é um avanço significativo, pois pode contribuir para diagnósticos precoces e intervenções clínicas mais eficazes. Entretanto, as dificuldades persistentes na correta identificação dos estágios mais avançados e intermediários indicam a necessidade de aprimoramento dos modelos, possivelmente através de técnicas de balanceamento de dados e otimização de hiperparâmetros. Em suma, este estudo evidencia o potencial das redes neurais profundas como ferramentas complementares no diagnóstico de demência, destacando a importância de abordagens metodológicas adaptativas para superar os desafios inerentes à classificação médica complexa. Embora o aprendizado por transferência reduza o custo computacional, o processo de ajuste das redes neurais ainda demanda considerável capacidade computacional e tempo para o treinamento, especialmente ao testar diferentes otimizadores e arquiteturas de modelos.

REFERÊNCIAS

1. ABADI M, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. In: USENIX Symposium on Operating Systems Design and Implementation (OSDI), 12., 2016, Savannah. Proceedings. Savannah: USENIX Association, 2016; 265-283.
2. ARCHANA B, KALIRAJAN K. Alzheimer's disease classification using convolutional neural networks. In: International Conference on Innovative Data Communication Technologies and Application (ICIDCA), 2023. IEEE, 2023; 1044-1048.
3. CASTELLANI RJ, et al. Alzheimer disease: Alzheimer's disease is the most common form of dementia. Disease-a-Month, 2010; 56(9): 484-546.
4. CHANDRA A, et al. Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment. Journal of Neurology, 2019; 266(6): 1293-1302.
5. EBRAHIMI A, et al. Multi-class classification of Alzheimer's disease using deep learning: a review. Current Alzheimer Research, 2021; 18(1): 50-68.
6. ESTEVA A, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017; 542(7639): 115-118.
7. GOODFELLOW I, et al. Deep learning. Cambridge: MIT Press, 2016; 800p.

8. GUZIOR N, et al. Alzheimer's disease: understanding the neuropsychological profile. Neuropsychology Review, 2015; 25(3): 1-15.
9. HE K, et al. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, Las Vegas. Proceedings. Los Alamitos: IEEE, 2016; 770-778.
10. HUANG G, et al. Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, Honolulu. Proceedings. Los Alamitos: IEEE, 2017; 4700-4708.
11. HUSSAIN MZ, et al. A tuned convolutional neural network model for accurate Alzheimer's disease classification. Scientific Reports, 2025; 15: 11616.
12. JACK CR Jr, et al. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology, 1997; 49(3): 786-794.
13. LA MONTAGNE P, et al. Medial temporal lobe subregional atrophy in aging and Alzheimer's disease: a longitudinal study. Frontiers in Aging Neuroscience, 2021; 13: 750154.
14. LIU M, et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. NeuroImage: Clinical, 2020; 27: 102303.
15. LITJENS G, et al. A survey on deep learning in medical image analysis. Medical Image Analysis, 2017; 42: 60-88.
16. MARCUS DS, et al. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience, 2007; 19(9): 1498-1507.
17. MISHRA RP, et al. Alzheimer's disease classification using convolutional neural network and random hyperparameter tuning. In: International Conference on Odisha on Electrical Power Engineering, Communication and Computing Technology (ODICON), 3., 2024, Bhubaneswar. Proceedings. Piscataway: IEEE, 2024; 1-6.
18. MORRIS JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology, 1993; 43(11): 2412-2414.
19. PASCANU R, et al. Understanding the difficulty of training deep feedforward neural networks. In: International Conference on Machine Learning (ICML), 30., 2013, Atlanta. Proceedings. Atlanta: JMLR, 2013; 1310-1318.
20. PENG L, et al. The role of demographic features in predicting Alzheimer's disease: a machine learning approach. Frontiers in Neuroscience, 2020; 14: 749.
21. RUBIN EH, et al. Mapping Mini-Mental State Examination scores onto Clinical Dementia Rating categories in Alzheimer's disease. American Journal of Geriatric Psychiatry, 1998; 6(2): 149-155.
22. SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition. arXiv 2014; 1409: 1556.
23. SZEGEDY C, et al. Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, Las Vegas. Proceedings. Los Alamitos: IEEE, 2016; 2818-2826.
24. TAN M, LE QV. EfficientNet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning (ICML), 36., 2019, Long Beach. Proceedings. Long Beach: JMLR, 2019; 6105-6114.
25. VAN DER MOLEN S. Scikit-learn for machine learning. Journal of Machine Learning Research, 2020; 21(1): 1-5.
26. ZHANG Z, et al. Evaluation metrics for machine learning in Alzheimer's disease: a review. Journal of Alzheimer's Disease, 2018; 62(3): 987-1002.
27. ZHU W, et al. Deep learning for medical image analysis: a comprehensive review. Medical Image Analysis, 2019; 58: 101552.