

Redes Neurais Convolucionais 3D para Classificação da Escala Fazekas em Ressonâncias Magnéticas Cerebrais

Lucas Junqueira Carvalho
DCC-UFMG
Belo Horizonte, Brasil
lucasjunqueiracarvalho@gmail.com

Júlio Guerra Domingues
DCC e FM-UFMG
Belo Horizonte, Brasil
juliodomingues@gmail.com

Leticia Ribeiro Miranda
DCC-UFMG
Belo Horizonte, Brasil
leticia-ribeiro98@hotmail.com

Indra Matsiendra Dias Ribeiro
DCC-UFMG
Belo Horizonte, Brasil
indramatsiendra@gmail.com

Etelvina Costa Sá Oliveira
DCC-UFMG
Belo Horizonte, Brasil
etelvina.oliveira2003@gmail.com

Wagner Meira Jr
DCC-UFMG
Belo Horizonte, Brasil
meira@dcc.ufmg.br

Gisele Lobo Pappa
DCC-UFMG
Belo Horizonte, Brasil
glpappa@dcc.ufmg.br

Bernardo de Mattos Viana
FM-UFMG
Belo Horizonte, Brasil
<https://orcid.org/0000-0001-8792-2996>

Debora Marques de Miranda
FM-UFMG
Belo Horizonte, Brasil
<https://orcid.org/0000-0002-7081-8401>

Marco Aurélio Romano Silva
FM-UFMG
Belo Horizonte, Brasil
<https://orcid.org/0000-0002-6558-4639>

Paula Ribeiro de Britto Borges
Grupo Fleury
Belo Horizonte, Brasil
paulabrittoborges@gmail.com

Luciana Costa-Silva
Grupo Fleury
Belo Horizonte, Brasil
lucianacosta@me.com

Abstract—Neurological disorders such as stroke and dementia increasingly affect aging populations, with White Matter Hyperintensities (WMHs) serving as important imaging biomarkers. The Fazekas scale, commonly used to grade WMHs in brain MRI, is limited by subjectivity and low reproducibility.

This study proposes an automated classification approach using 3D brain MRI and deep convolutional neural networks (DenseNet-121 and DenseNet-169). Both models were adapted to process volumetric data and trained on a clinically acquired, radiologist-labeled dataset, following standard preprocessing and augmentation.

The 3D DenseNet-121 achieved 89.8% accuracy and 89.2% F1-score in distinguishing mild (Fazekas 0/1) from moderate/severe cases (Fazekas 2/3), while DenseNet-169 reached 87.5% accuracy and 85.3% F1-score. Results in multiclass settings were also competitive.

These findings support the use of deep learning for reliable and consistent automated Fazekas grading, contributing to improved neurological diagnosis and early risk assessment.

Index Terms—White matter hyperintensities, Fazekas scale, brain MRI, deep learning, convolutional neural networks, DenseNet, medical image analysis.

I. INTRODUÇÃO

As doenças neurológicas representam uma das principais causas de incapacidade e morte em todo o mundo, com impacto crescente em sociedades envelhecidas. Diagnosticar precocemente essas condições é um desafio crítico, pois muitas

lesões cerebrais se desenvolvem de forma silenciosa e gradual, sendo detectáveis apenas por meio de exames de imagem como a ressonância magnética (RM). No entanto, interpretar essas imagens com precisão exige tempo, expertise médica e está sujeito a variações interobservadores.

Nesse contexto, destacam-se as hiperintensidades da substância branca (WMHs) — áreas anormalmente brilhantes observadas principalmente nas sequências T2 e FLAIR da RM, situadas na substância branca profunda e subcortical [1]. Embora possam surgir como parte do envelhecimento natural, as WMHs estão fortemente associadas a doenças cerebrovasculares, hipertensão crônica, declínio cognitivo e demência vascular [1].

Para descrever a extensão e gravidade dessas lesões, utiliza-se a Escala de Fazekas, uma classificação visual proposta em 1987 [2]. Essa escala avalia separadamente as lesões periventriculares e as profundas, atribuindo escores de 0 a 3 com base em sua aparência e extensão. Apesar de amplamente utilizada na prática clínica, essa avaliação é subjetiva, o que motiva o desenvolvimento de métodos automatizados de análise.

Ademais, as WMHs são reconhecidas como importantes biomarcadores de lesão cerebral subclínica, estando associadas a maior risco de acidente vascular cerebral, piora cognitiva e prognóstico clínico em pacientes com doenças crônicas como hipertensão e diabetes [3]. Estudos indicam que, embora sejam

raras em adultos jovens, as WMHs podem ser detectadas em até 90% dos indivíduos com menos de 65 anos, sendo que sua carga basal pode prever o surgimento de demência décadas depois [3].

Diante deste cenário, um dos principais desafios na predição da pontuação de Fazekas está na natureza subjetiva da avaliação visual, que pode variar entre especialistas e está suscetível a inconsistências. Além disso, a heterogeneidade das WMHs, em termos de localização, intensidade e padrão de distribuição, dificulta a classificação. A complexidade aumenta com a sobreposição de sinais com outras alterações cerebrais, como atrofas e lesões isquêmicas, exigindo abordagens de aprendizado de máquina que consigam capturar variações espaciais e contextuais das imagens de ressonância magnética. Neste artigo, propomos, implementamos e analisamos um modelo para avaliar a carga de lesões da substância branca a partir de imagens de ressonância magnética cerebral. Para isso, empregamos uma arquitetura de rede neural profunda do tipo Convolutional Neural Network (CNN), capaz de extrair e aprender padrões, a fim de prever os escores da escala de Fazekas.

Este trabalho está organizado da seguinte forma: Na primeira parte, são mencionados as pesquisas que atuaram no estudo de WMHs e a Escala Fazekas. Em seguida, descreve-se o processo de aquisição dos dados utilizados nesta pesquisa. Posteriormente, é detalhada a metodologia utilizada nas análises, abrangendo o pré-processamento e aumento artificial dos dados, a arquitetura da rede neural implementada, sua adaptação ao domínio tridimensional e o processo de classificação. Na sequência, são discutidos os resultados obtidos nos experimentos. Por fim, são apresentadas as discussões e conclusões do trabalho, além das perspectivas para trabalhos futuros e os agradecimentos.

II. TRABALHOS RELACIONADOS

Nas últimas décadas, o avanço das técnicas de aprendizado de máquina tem impulsionado significativamente a análise de imagens médicas, com destaque para a utilização de redes neurais profundas em tarefas como segmentação e classificação de lesões cerebrais. Em particular, diversos estudos têm aplicado essas abordagens para analisar as hiperintensidades da substância branca (WMHs) em exames de ressonância magnética, utilizando modelos convolucionais e arquiteturas do tipo U-Net para segmentar automaticamente essas regiões e estimar escores clínicos relevantes. A escala de Fazekas, que quantifica a gravidade das WMHs, tem sido frequentemente utilizada como referência em tais estudos, dada sua forte associação com piores desfechos clínicos, como maior incidência de AVC recorrente, declínio cognitivo e redução da mobilidade em idosos. Ao combinar a expertise médica com métodos computacionais avançados, esses trabalhos têm buscado melhorar a acurácia e eficiência na detecção e classificação de lesões, contribuindo para a identificação precoce de indivíduos com maior risco de comprometimento neurológico.

No estudo [4], desenvolveu-se um modelo que tinha como objetivo prever as avaliações na escala Fazekas e identificar regiões no cérebro com alto risco de mudanças na hiperintensidade da substância branca ao usar técnicas de aprendizado profundo, juntamente com T2-FLAIR. O modelo classificou com precisão a gravidade das hiperintensidades da substância branca (WMH), com acurácia de 0.656, precisão de 0.6296, revocação de 0.5937 e F1-score de 0.5768 no conjunto de treinamento, e acurácia de 0.621, precisão de 0.6633, recall de 0.6120 e F1-score de 0.6018 em dois conjuntos independentes. Além disso, ele localizou regiões de alto risco que corresponderam a alterações estruturais e funcionais no cérebro ao longo do tempo. A análise por ressonância magnética multimodal revelou que essas regiões de alto risco estavam associadas à atrofia da substância cinzenta e branca, danos microestruturais, perfusão reduzida e declínio cognitivo, sugerindo o potencial do modelo para a detecção precoce de alterações cerebrais relacionadas às WMHs e comprometimento cognitivo.

Já o estudo [5] comparou os danos microvasculares cerebrais na doença de Alzheimer (DA) e na tauopatia primária relacionada à idade (PART) utilizando a escala de Fazekas em exames de ressonância magnética. Observou-se que os escores de Fazekas se correlacionaram de forma distinta à atrofia cerebral regional nos dois grupos. Na DA, as correlações foram modestas, com coeficientes de Pearson entre 0.18 e 0.30 nas regiões temporal medial, temporal anterior e fronto-insular. Na PART, as correlações foram mais fortes, especialmente na região fronto-insular, onde os valores de R alcançaram 0.53 (periventricular) e 0.62 (profundo). Esses achados sugerem que a PART possui uma patologia microvascular distinta da DA, reforçando seu status como uma condição neurodegenerativa separada.

O trabalho de [6] se destaca ao apresentar e validar um modelo de aprendizado profundo para classificar os graus da escala de Fazekas e diferenciar demência vascular subcortical. Para isso, os autores utilizaram redes neurais convolucionais (CNN) aplicadas a imagens de ressonância magnética FLAIR de 800 pacientes, obtidas em um hospital terciário. O modelo foi treinado com uma arquitetura baseada em U-Net e demonstrou um bom desempenho, com valores elevados da área sob a curva (AUC), variando de 0.92 a 0.96 na classificação dos diferentes graus da escala de Fazekas. Os valores de corte ideais para categorizar o volume de hiperintensidades da substância branca (WMH) foram 3.4 mL (normal vs. leve/moderado/grave), 9.6 mL (normal/leve vs. moderado/grave) e 17.1 mL (normal/leve/moderado vs. grave), com AUCs correspondentes de 0.921, 0.956 e 0.960. Ao diferenciar normal/leve de moderado/grave no conjunto de teste, o modelo atingiu sensibilidade de 96.4%, especificidade de 89.9% e acurácia de 91.7%. Já na distinção entre demência vascular subcortical e outras condições, os valores de sensibilidade, especificidade e acurácia foram de 83.3%, 84.3% e 84.3%, respectivamente.

Outra pesquisa relevante, realizado por [7], destaca-se ao explorar os padrões espaciais das hiperintensidades da substância branca (WMH) e suas associações com fatores vasculares

e biomarcadores de Alzheimer (amiloide e tau). Diferente de outras pesquisas que enfatizam a ligação entre WMH e patologia de Alzheimer, este trabalho encontrou apenas uma associação fraca entre WMH e amiloidose, com coeficiente de correlação R igual a 0.07 e valor de p inferior a 0.001, e nenhuma associação com a carga de tau, sugerindo que os mecanismos vasculares têm um papel mais relevante na presença de WMH do que os processos neurodegenerativos típicos da doença de Alzheimer. Um ponto importante foi a ausência de diferenças significativas entre WMH periventriculares e profundas em relação à cognição e aos fatores de risco, o que reforça a ideia de que ambas compartilham mecanismos subjacentes comuns. As medições de WMH periventricular e profunda apresentaram correlações semelhantes envolvendo idade, escore de condições cardiometabólicas, que refletem o risco vascular, e cognição, todas com valor de p inferior a 0.001, reforçando a hipótese de que ambas compartilham mecanismos subjacentes comuns. Além disso, a boa reprodutibilidade entre diferentes equipamentos de ressonância e a forte concordância entre as escalas visuais e os volumes automatizados de WMH tornam esses achados relevantes para futuras pesquisas multicêntricas, possibilitando maior confiabilidade na comparação de dados obtidos em diferentes contextos clínicos.

Esses estudos evidenciam o papel crescente das técnicas de aprendizado de máquina na análise de imagens médicas, especialmente na classificação e quantificação das hiperintensidades da substância branca segundo a escala Fazekas. Eles ressaltam tanto o potencial dessas abordagens para auxiliar na identificação precoce de alterações cerebrais e risco clínico, quanto os desafios associados à interpretação dos resultados e à variabilidade dos dados entre diferentes populações e contextos clínicos. O nosso trabalho contribui para essa linha de pesquisa ao desenvolver modelos baseados em redes neurais convolucionais (CNN) para prever os escores de Fazekas a partir de exames de ressonância magnética volumétricos.

Esses estudos evidenciam o papel crescente das técnicas de aprendizado de máquina na análise de imagens médicas, especialmente na classificação e quantificação das hiperintensidades da substância branca segundo a escala de Fazekas. Eles ressaltam tanto o potencial dessas abordagens para auxiliar na identificação precoce de alterações cerebrais e risco clínico, quanto os desafios associados à interpretação dos resultados e à variabilidade dos dados entre diferentes populações e contextos clínicos.

O presente trabalho se diferencia dos trabalhos anteriores ao utilizar como entrada séries completas de exames de ressonância magnética volumétrica, ao invés de fatias específicas ou regiões selecionadas, permitindo que o modelo aprenda representações tridimensionais mais abrangentes. Essa abordagem possibilitou que a DenseNet-121 adaptada atingisse uma acurácia de 89.8% e F1-score de 89.2% na classificação binária (Fazekas 0/1 vs. 2/3), superando os resultados reportados por [4] (acurácia de 62.1% e F1-score de 60.2%), e apresentando desempenho comparável ao modelo proposto por [6], que utilizou segmentações manuais e arquitetura baseada em

U-Net, na tarefa de classificação binária (Fazekas 0/1 vs. 2/3). Outro diferencial relevante deste trabalho está na versatilidade dos experimentos realizados, incluindo (i) classificação binária (Fazekas 0/1 vs. 2/3), (ii) classificação tri-classe (Fazekas 0 vs. 1 vs. 2/3) e (iii) classificação multiclasse (Fazekas 0 vs. 1 vs. 2 vs. 3), o que amplia a aplicabilidade clínica do modelo e demonstra sua robustez frente a diferentes cenários diagnósticos.

III. METODOLOGIA

A fim de investigar a eficácia de redes neurais convolucionais profundas na classificação automática de exames MRI segundo a escala Fazekas, foi desenvolvida uma metodologia composta por etapas sucessivas de pré-processamento, aumento artificial de dados, adaptação arquitetural e treinamento supervisionado. Inicialmente, os volumes de imagem foram normalizados e padronizados em dimensão para garantir consistência entre amostras. Em seguida, aplicaram-se técnicas de enriquecimento de dados (*data augmentation*), com o objetivo de expandir a diversidade do conjunto de treinamento e reduzir o risco de sobreajuste. Posteriormente, utilizou-se as arquiteturas DenseNet-169 e DenseNet-121 [8], adaptadas do domínio bidimensional para o tridimensional, com o intuito de explorar a continuidade volumétrica dos dados médicos. Por fim, definiu-se um protocolo de treinamento voltado para diferentes cenários de classificação conforme os graus da escala Fazekas. As subseções a seguir detalham cada uma dessas etapas.

A. Aquisição dos Dados

Os dados utilizados neste estudo consistem em uma coleção retrospectiva de 10000 exames de RM do encéfalo, provenientes de um serviço de diagnóstico por imagem em Minas Gerais, Brasil. Os exames foram adquiridos em equipamentos de três diferentes fabricantes: Siemens (MAGNETOM Avanto 1,5T e MAGNETOM Skyra 3,0T), Philips (Ingenia 1,5T e Prodiva CS 1,5T) e GE Healthcare (Signa HDx 1,5T e Signa HDxt 1,5T). Independentemente do equipamento de origem, foram selecionadas para análise as imagens da sequência axial T2-FLAIR (Fluid-Attenuated Inversion Recovery), por ser a mais indicada para a visualização de hiperintensidades da substância branca, e seus respectivos laudos radiológicos em texto livre. O objetivo metodológico foi extrair e validar a pontuação da escala Fazekas, registrada de forma não estruturada nos laudos.

1) *Aquisição dos Rótulos*: O processo de rotulagem iniciou-se com o pré-processamento dos laudos textuais, que incluiu a conversão para minúsculas, remoção de stopwords e segmentação estrutural. Em seguida, adotou-se uma estratégia híbrida para a extração da pontuação. Primeiramente, uma abordagem baseada em Expressões Regulares (Regex) foi aplicada para identificar menções explícitas ao termo "fazekas". Para ampliar a cobertura, especialmente de casos normais (Fazekas 0), uma segunda abordagem de descoberta de padrões foi implementada a partir de um subconjunto de 50 laudos de referência, previamente anotados por especialistas. Nesta fase, foram empregadas duas técnicas em paralelo: (i) a extração de

n-gramas (n=2 a 7) com seleção dos mais relevantes via teste qui-quadrado (χ^2) [12], [13], e (ii) a sumarização extrativa com um Large Language Model (LLM, Google Gemini 1.5) para identificar sentenças-chave [14]. As regras e padrões gerados foram curados por radiologistas e aplicados aos laudos restantes. Finalmente, para garantir a precisão dos rótulos, foi realizada uma curadoria por especialistas: laudos com anotações ambíguas foram revisados e, adicionalmente, todos os exames classificados como Fazekas 2 e 3, por sua menor prevalência, foram submetidos a uma dupla avaliação por dois profissionais com atuação em neurorradiologia, com a classificação final definida por consenso. Ao final de todo o processo, consolidou-se um dataset de 390 exames com rótulos da escala Fazekas validados, sendo eles classificados da seguinte forma: 94 exames com escore 0, 191 com escore 1, 56 com escore 2 e 49 com escore 3.

B. Pré-Processamento dos Dados

Inicialmente, cada imagem das séries de exame MRI foi normalizada com o método min-max de modo a restringir os valores de intensidade ao intervalo $[0, 1]$:

$$I_{norm}(x, y) = \frac{I(x, y) - I_{min}}{I_{max} - I_{min}}, \quad (1)$$

onde $I(x, y)$ representa a intensidade original de cada pixel e I_{min} , I_{max} são os valores mínimo e máximo da imagem.

Além disso, para garantir uniformidade espacial, todas as imagens são redimensionadas para 256×256 pixels por meio de interpolação bilinear. Para assegurar consistência temporal entre os volumes, o número de fatias D é uniformizado para um valor fixo de 30. Volumes com menos de 30 fatias são preenchidos com *zero-padding* ao final da dimensão de profundidade, enquanto volumes com mais de 30 fatias são truncados para as primeiras 30:

$$V' = \begin{cases} \text{pad}(V, 30 - D), & \text{se } D < 30 \\ V[:, : 30, :, :], & \text{se } D > 30 \end{cases}$$

Essa uniformização permite que todos os volumes possuam o formato final (1, 30, 256, 256).

C. Aumento Artificial de Dados

Durante o treinamento do modelo, foram aplicadas técnicas de data augmentation com o objetivo de aumentar a variabilidade das amostras e mitigar o risco de sobreajuste. Três transformações principais foram utilizadas:

Primeiramente, aplicou-se o espelhamento horizontal [9], realizado por meio da reflexão do volume ao longo do eixo correspondente à largura da imagem (eixo W). Essa operação consiste em inverter os valores da última dimensão do tensor, promovendo simetria espacial nos dados. Em seguida, foram empregadas rotações nos planos axiais das imagens [9], com ângulos de 90° , 180° e 270° . Essas rotações foram implementadas através de combinações de transposições e reflexões dos eixos espaciais dos tensores, simulando variações na orientação anatômica. Por fim, utilizou-se a deformação

elástica [10], uma técnica que aplica deslocamentos não lineares suaves ao volume. Esses deslocamentos são gerados a partir de campos vetoriais aleatórios suavizados com filtros gaussianos (com parâmetros $\alpha = 15$ e $\sigma = 3$), conforme a metodologia proposta em [11]. Essa transformação visa simular pequenas variações morfológicas naturais, preservando a coerência das estruturas anatômicas presentes nas imagens.

D. DenseNet

As *Densely Connected Convolutional Networks* (DenseNets) [8] são uma classe de redes neurais convolucionais caracterizadas pelo uso de conexões densas entre camadas. Ao contrário das arquiteturas tradicionais, em que cada camada recebe entrada apenas da camada imediatamente anterior, as DenseNets estabelecem conexões diretas de todas as camadas anteriores para todas as camadas subsequentes dentro de um mesmo bloco denso (*dense block*).

Seja $\mathbf{x}_0$ a entrada inicial de um bloco denso, e $\mathbf{x}_\ell$ a saída da ℓ -ésima camada. A operação de cada camada pode ser definida por

$$\mathbf{x}_\ell = H_\ell([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{\ell-1}]), \quad (2)$$

onde $[\cdot]$ representa a concatenação ao longo do eixo dos canais e $H_\ell(\cdot)$ é uma função composta por operações de convolução, normalização em lote (batch normalization) e ativação ReLU, a saber:

$$H_\ell(\cdot) = \text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3}(\cdot))). \quad (3)$$

Para evitar o crescimento exponencial do número de canais, a DenseNet utiliza um parâmetro chamado *growth rate*, denotado por k , que define o número de novos mapas de características produzidos por camada. Assim, após ℓ camadas, o número total de canais é $F_0 + k \cdot \ell$, onde F_0 é o número de canais da entrada inicial do bloco.

Entre blocos densos, utiliza-se uma camada de transição que aplica convolução 1×1 seguida de redução de dimensionalidade via *average pooling*:

$$\mathbf{x}_{trans} = \text{AvgPool}(\text{BN}(\text{Conv}_{1 \times 1}([\mathbf{x}_0, \dots, \mathbf{x}_\ell]))). \quad (4)$$

As equações (2) a (4) descrevem o funcionamento fundamental da DenseNet. Tal arquitetura apresenta vantagens notáveis em termos de eficiência de parâmetros e desempenho.

A arquitetura DenseNet tem sido amplamente utilizada em aplicações de imagem médica devido à sua capacidade de extrair representações discriminativas [12], uma característica comum nesse domínio. Além disso, a estrutura compacta da DenseNet, com menor número de parâmetros em comparação com arquiteturas profundas tradicionais (como VGG ou ResNet), é particularmente vantajosa em dispositivos médicos embarcados ou sistemas de inferência com restrições computacionais.

E. Adaptação para dados 3D

A DenseNet é originalmente projetada para imagens bidimensionais (2D), sendo amplamente utilizada em tarefas de classificação de imagens naturais. No entanto, como exames de ressonância magnética (MRI) são volumétricos, foi feita

a conversão da DenseNet para o âmbito 3D para explorar as características espaciais *interslice* presentes nos dados.

O processo de adaptação consiste em converter sistematicamente todos os módulos convolucionais, de normalização e de pooling da arquitetura original de 2D para 3D, preservando os pesos pré-treinados no ImageNet. O procedimento é detalhado a seguir:

- 1) **Conversão de Camadas Convolucionais:** Cada camada `Conv2d` é convertida para uma `Conv3d` com kernel tridimensional do tipo $(1, k, k)$, mantendo os filtros originais nas dimensões espaciais e introduzindo um fator neutro no eixo de profundidade. Os pesos 2D são replicados ao longo da nova dimensão de profundidade:

$$W_{3D}(c_{out}, c_{in}, 0, i, j) = \frac{W_{2D}(c_{out}, c_{in}, i, j)}{k} \quad \forall c_{in}, c_{out}, i, j \quad (5)$$

onde W_{2D} representa os pesos da convolução 2D original e k é o tamanho do kernel.

- 2) **Conversão de Normalizações e Funções de Ativação:** Camadas `BatchNorm2d` são convertidas para `BatchNorm3d`, mantendo o número de canais. As ativações `ReLU` são preservadas com suas configurações *inplace*.
- 3) **Conversão de Operações de Pooling:** Camadas de pooling 2D são adaptadas para 3D com kernel $(1, k, k)$, permitindo redução espacial sem compressão da profundidade volumétrica. Assim, mantém-se a resolução entre fatias adjacentes, o que é fundamental para preservar a coerência anatômica ao longo do volume.
- 4) **Ajuste da Primeira Camada:** A camada inicial de convolução, originalmente projetada para três canais RGB, é substituída por uma camada com número arbitrário de canais de entrada C_{in} , tipicamente $C_{in} = 1$ para imagens médicas em tons de cinza.
- 5) **Classificador:** Após as camadas de extração de características, utiliza-se uma operação de pooling global adaptativo em três dimensões (`AdaptiveAvgPool3d`) para reduzir a saída a um vetor de dimensão fixa, que é então passado por uma camada linear para realizar a classificação:

$$y = \mathbf{W} \cdot \text{flatten}(\text{AvgPool3D}(f(\mathbf{x}))) + \mathbf{b}. \quad (6)$$

- 6) **Reutilização de Pesos:** Os pesos do modelo 2D pré-treinado no ImageNet são carregados, garantindo uma inicialização mais informada e potencialmente melhor desempenho em regime de poucos dados (*transfer learning*). A replicação de filtros nas dimensões volumétricas permite essa transferência parcial, mesmo em domínios diferentes.

A adaptação descrita permite que a DenseNet explore tanto as correlações espaciais bidimensionais quanto a continuidade volumétrica dos dados médicos, tornando-a especialmente adequada para a tarefa de classificação de exames MRI.

F. Tarefa de Classificação

A escala de Fazekas é um importante indicador de fatores de risco neurológicos, especialmente nos graus 2 e 3, frequentemente associados a diversas doenças. Para investigar a capacidade de redes neurais convolucionais em identificar esses fatores de risco a partir de imagens de ressonância magnética, este trabalho propõe a avaliação de modelos em três cenários distintos de classificação: (i) binária — Fazekas 0/1 vs. 2/3, (ii) triclasse — Fazekas 0 vs. 1 vs. 2/3, e (iii) multiclasse — Fazekas 0 vs. 1 vs. 2 vs. 3.

G. Configuração de Treino

Para a DenseNet-169, foi utilizado o otimizador ADAM com taxa de aprendizado 1×10^{-5} , tamanho de *batch* 3, devido as restrições de recursos computacionais e 10 épocas de treinamento. Para a DenseNet-121, foi utilizado o otimizador ADAM com taxa de aprendizado 1×10^{-5} , tamanho de *batch* de 4, devido as restrições de recursos computacionais e 10 épocas de treinamento. Além disso, os conjuntos de treino, validação e teste receberam amostras que correspondem a, respectivamente, 70%, 15% e 15% dos dados do dataset completo, sendo que, foram utilizadas sementes aleatórias fixas durante esta amostragem para fins de replicação de experimentos e comparação de resultados. Por fim, ambas as redes foram treinadas com a GPU NVIDIA GeForce RTX 4090 (24 GB).

IV. RESULTADOS

Os resultados da classificação dos arquivos MRI na escala Fazekas estão descritos em "Table I". Considerando os resultados relacionados a rede Densenet-169 e seus respectivos intervalos de confiança, para a diferenciação de Fazekas 0/1 vs. 2/3, acurácia, precisão, revocação e F1-score foram 87.5%(CI:86.4%-89.8%), 85.3%(CI:82.4%-87.3%), 86.5%(CI:86.4%-86.7%) e 85.3%(CI:84.3%-86.2%), respectivamente. Para a diferenciação de Fazekas 0 vs. 1 vs. 2/3, acurácia, precisão, revocação e F1-score foram 85.2%(CI:79.6%-91.5%), 86.3%(CI:81.4%-91.4%), 85.2%(CI:79.6%-91.5%) e 85.3%(CI:84.3%-86.2%), respectivamente. Por fim, para a diferenciação de Fazekas 0 vs. 1 vs. 2 vs. 3, acurácia, precisão, revocação e F1-score foram 83%(CI:76.2%-91.6%), 84.2%(CI:79.7%-92%), 83%(CI:76.2%-91.6%) e 83.3%(CI:75.8%-91.6%), respectivamente.

Considerando os resultados relacionados a rede Densenet-121 e seus respectivos intervalos de confiança, para a diferenciação de Fazekas 0/1 vs. 2/3, acurácia, precisão, revocação e F1-score foram 89.8%(CI:86.4%-94.9%), 90.6%(CI:88%-95.2%), 89.8%(CI:86.4%-94.9%) e 89.2%(CI:85.4%-94.5%), respectivamente. Para a diferenciação de Fazekas 0 vs. 1 vs. 2/3, acurácia, precisão, revocação e F1-score foram 83%(CI:79.6%-88.1%), 83.9%(CI:79.4%-89.6%), 83%(CI:79.6%-88.1%) e 82.7%(CI:79.4%-87.3%), respectivamente. Por fim, para a diferenciação de Fazekas 0 vs. 1 vs. 2 vs. 3, acurácia, precisão, revocação e

F1-score foram 79.6%(CI:77.9%-81.3%), 82.3%(CI:82.1%-82.6%), 79.6%(CI:77.9%-81.3%) e 79.3%(CI:77.2%-81.4%), respectivamente.

V. DISCUSSÃO

A escala fazekas pode ser uma ferramenta útil na área da medicina para, ao quantificar alterações na substância branca do cérebro, determinar os fatores de risco em um paciente. Fazekas 1 pode representar alterações vasculares leves relacionadas ao envelhecimento normal, especialmente em indivíduos idosos sem doenças neurológicas conhecidas, por outro lado, fazekas 2 e 3 representam alterações mais graves na massa encefálica e podem ser indicadores de doenças mais graves, por isso, a identificação destas escalas é relevante.

Neste trabalho foram utilizadas redes neurais convolucionais para, dado uma série de imagens volumétrica MRI, classificar automaticamente a escala Fazekas com bons resultados empíricos. Foi possível realizar a diferenciação de Fazekas 0/1 vs. 2/3 com as métricas de acurácia, precisão, revocação e F1-score como 89.8%(CI:86.4%-94.9%), 90.6%(CI:88%-95.2%), 89.8%(CI:86.4%-94.9%) e 89.2%(CI:85.4%-94.5%), respectivamente. Para realizar a diferenciação de Fazekas 0 vs. 1 vs. 2/3, foi possível obter as métricas de acurácia, precisão, revocação e F1-score como 85.2%(CI:79.6%-91.5%), 86.3%(CI:81.4%-91.4%), 85.2%(CI:79.6%-91.5%) e 85.2%(CI:79.6%-91.4%), respectivamente. E, também, Foi possível realizar a diferenciação de fazekas 0 vs. 1 vs. 2 vs. 3 com as métricas de acurácia, precisão, revocação e F1-score como 83.6%(CI:76.2%-91.6%), 85.3%(CI:79.7%-92%), 83.6%(CI:76.2%-91.6%) e 83.5%(CI:75.8%-91.6%), respectivamente.

Continuando, fazendo uma comparação dos resultados obtidos neste trabalho com os trabalhos relacionados, foi possível, com menos pré-processamento das imagens, obter resultados superiores aos obtidos em [4]. Além disso, foi possível obter resultados comparáveis a [6] na tarefa de classificação de Fazekas 0/1 vs. 2/3, e obter resultados adicionais com modalidades de classificação Fazekas que não foram exploradas em [6]. Por fim, este trabalho pode explorar a análise do desempenho das redes convolucionais profundas na classificação de dados volumétricos, o que é um diferencial dos trabalhos relacionados, que tiveram de selecionar fatias específicas dos exames MRI para usar de entrada para estas redes convolucionais.

VI. CONCLUSÕES

Este trabalho demonstrou que o uso de redes neurais convolucionais pode ser uma boa ferramenta na área da medicina para, ao fazer a classificação automática de exames MRI na escala Fazekas, medir fatores de risco de um paciente. Em especial evidenciou-se o potencial das CNNs adaptadas para o 3D na tarefa de classificação da carga de lesões da substância branca, segundo a escala de Fazekas, a partir de exames de ressonância magnética cerebral. A aplicação de arquiteturas DenseNet-121 e DenseNet-169, juntamente com as técnicas

de pré-processamento e aumento artificial de dados, permitiu alcançar bom resultados.

Os modelos foram capazes de diferenciar adequadamente entre os níveis da escala, com destaque para a distinção entre grupos de menor e maior gravidade (Fazekas 0/1 vs. 2/3), atingindo acurácia de até 89.8%(CI:86.4%-94.9%) e F1-score de 89.2%(CI:85.4%-94.5%). Esses resultados reforçam a viabilidade do uso de modelos de aprendizado profundo como ferramentas auxiliares no suporte ao diagnóstico neurológico, especialmente em ambientes clínicos onde a avaliação subjetiva pode ser limitada por variabilidade dos especialistas.

Além disso, os resultados obtidos neste trabalho estão alinhados com os avanços recentes da literatura, contribuindo com análise baseada em imagens volumétricas e reforçando a importância da adaptação arquitetural ao contexto tridimensional. A boa performance alcançada, mesmo com conjuntos de dados relativamente pequenos, ressalta a eficiência das DenseNets e o resultado da transferência de aprendizado (transfer learning) a partir de redes pré-treinadas.

Por fim, os resultados apontam para a aplicabilidade de modelos baseados em deep learning na análise de biomarcadores de lesão cerebral subclínica, com capacidade para apoiar a detecção precoce de alterações cerebrais e no acompanhamento de pacientes com fatores de risco para demência e doenças cerebrovasculares.

VII. TRABALHOS FUTUROS

Um desafio encontrado na classificação de exames MRI na escala fazekas utilizando dados médicos reais é a variabilidade contida nos dados. Um time de médicas pode classificar um exame de um modo, enquanto outro time de médicos pode ter uma opção de classificação distinta, o que pode atrapalhar o desempenho de uma rede neural para encontrar padrões durante o treinamento. Portanto, pode ser possível, com o apoio de radiologistas, realizar a validação dos dados para permitir um melhor desempenho na tarefa de classificação.

Além disso, diferentemente das redes convolucionais, que operam localmente e dependem de camadas empilhadas para capturar dependências globais, redes baseadas em transformers utilizam mecanismos de atenção global, permitindo que cada fatia do volume de um exame MRI interaja diretamente com todas as outras, independentemente da distância entre elas. Essa característica pode ser particularmente útil na classificação de imagens cerebrais, em que informações clínicas relevantes podem estar distribuídas de forma difusa ou sutil ao longo do volume. Dessa forma, pode ser possível utilizar uma rede transformers para se obter um desempenho superior ao encontrado pelas redes convolucionais neste trabalho.

AGRADECIMENTOS

Este projeto foi parcialmente financiado pelo CNPq, CAPES, FAPEMIG, INCTNeurotecR e IAIA - INCT em IA.

TABLE I
DESEMPENHO NAS TAREFAS DE CLASSIFICAÇÃO DA ESCALA DE FAZEKAS

Tarefa de Classificação	Rede	Acurácia	Precisão	Revocação	F1-score
Fazekas 0/1 vs. 2/3	DenseNet-121	89.8%(86.4%-94.9%)	90.6%(88%-95.2%)	89.8%(86.4%-94.9%)	89.2%(85.4%-94.5%)
	DenseNet-169	87.5%(86.4%-89.8%)	85.3%(82.4%-87.3%)	86.5%(86.4%-86.7%)	85.3%(84.3%-86.2%)
Fazekas 0 vs. 1 vs. 2/3	DenseNet-121	83%(79.6%-88.1%)	83.9%(79.4%-89.6%)	83%(79.6%-88.1%)	82.7%(79.4%-87.3%)
	DenseNet-169	85.2%(79.6%-91.5%)	86.3%(81.4%-91.4%)	85.2%(79.6%-91.5%)	85.2%(79.6%-91.4%)
Fazekas 0 vs. 1 vs. 2 vs. 3	DenseNet-121	79.6%(77.9%-81.3%)	82.3%(82.1%-82.6%)	79.6%(77.9%-81.3%)	79.3%(77.2%-81.4%)
	DenseNet-169	83.6%(76.2%-91.6%)	85.3%(79.7%-92%)	83.6%(76.2%-91.6%)	83.5%(75.8%-91.6%)

REFERENCES

- [1] Schmid, S., Shamaei, A., Souza, R., & Frayne, R. "Quantifying White Matter Hyperintensities: Predicting Periventricular Fazekas Scores with Uncertainty Estimation." 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). IEEE, 2025.
- [2] F. Fazekas, J. B. Chawluk, A. Alavi, et al., "MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging," *AJR Am J Roentgenol*, vol. 149, no. 2, pp. 351-356, Aug. 1987.
- [3] d'Arbeloff, T., et al. "White matter hyperintensities are common in midlife and already associated with cognitive decline." *Brain Communications*, v. 1, n. 1, p. fcz041, 2019.
- [4] Si Mu, Weizhao Lu, Guanghui Yu, Lei Zheng, Jianfeng Qiu, "Deep learning-based grading of white matter hyperintensities enables identification of potential markers in multi-sequence MRI data", *Computer Methods and Programs in Biomedicine*, vol. 243, January 2024.
- [5] Quintas-Neves, M., Almeida, F.C., Gauthreaux, K. et al. Fazekas scale magnetic resonance imaging assessment in Alzheimer's disease and primary age-related tauopathy. *Neuroradiology* 66, 2185–2193 (2024).
- [6] Joo L., Shim, W. H., Suh, C. H., et al. "Diagnostic performance of deep learning-based automatic white matter hyperintensity segmentation for classification of the Fazekas scale and differentiation of subcortical vascular dementia". *PLoS One* 17, e0274562 (2022).
- [7] Pradeep, A., Raghavan, S., Przybelski, S. A., et al. "Can white matter hyperintensities based Fazekas visual assessment scales inform about Alzheimer's disease pathology in the population?." *Alzheimer's Research & Therapy* 16.1 (2024): 157.
- [8] G. Huang, Z. Liu, L. van der Maaten, and K. Weinberger, "Densely connected convolutional networks," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017.
- [9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM* 60, 6 (June 2017), 84–90.
- [10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015; pp 234–241.
- [11] J.-S. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," *Proc. Int. Conf. Document Analysis and Recognition (ICDAR)*, pp. 958–963, 2003.
- [12] T. Zhou, X.Y. Ye, H.L. Lu, "Dense convolutional network and its application in medical image analysis," *Proc. Int. Conf. Document Analysis and Recognition (ICDAR)*, *BioMed Res. Int.*, 2022 (2022), p. 22.
- [13] John S. Malamon, "DNA N-gram analysis framework (DNAnamer): A generalized N-gram frequency analysis framework for the supervised classification of DNA sequences," *Heliyon*, Volume 10, Issue 17, 2024.
- [14] Marcio Sarmiento and Hilário de Oliveira, "Sumarização Automática de Artigos de Notícias em Português: Da Extração à Abstração com Abordagens Clássicas e Modelos de Neurais," *Anais do XV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pp. 139-148, 2024.