Classificação Binária de Imagens de Ressonância Magnética de Osteoartrite com o Modelo R3D_18 Modificado

Thalles C. Fontainha[1], **Felipe da R. Henriques**[1,2], **Amaro A. Lima**[1,3], **Gabriel M. Araujo**[4], **Ricardo de S. Tesch**[5]

[1]Programa de Pós-Graduação em Instrumentação e Óptica Aplicada – CEFET/RJ
[2]Programa de Pós-Graduação em Ciência da Computação – CEFET/RJ
[3]Departamento de Telecomunicações – CEFET/RJ, Campus Nova Iguaçu
[4]Programa de Pós-Graduação em Engenharia Elétrica – CEFET/RJ
[5]Departamento de Medicina Regenerativa – UNIFASE

thalles.fontainha@aluno.cefet-rj.br,
{felipe.henriques, amaro.lima, gabriel.araujo}@cefet-rj.br,
ricardotesch@prof.unifase-rj.edu.br

Abstract. *This work proposes a 3D convolutional neural network model derived from the R3D_18 architecture to classify magnetic resonance images (MRI) of the knee, differentiating normal and abnormal images associated with osteoarthritis. Using the OAI-MRI-3DDESS dataset, the model was trained with cross-validation and oversampling to compensate for class imbalance. The approach aims to improve diagnostic accuracy and reduce professionals' workload by exploring the volumetric information of the images. The experiments indicated robust performance, with accuracy higher than 86% and AUC close to 0.92, showing the method's potential for clinical applications.*

Resumo. *Este trabalho propõe um modelo de rede neural convolucional 3D derivado da arquitetura R3D_18 para classificar imagens de ressonância magnética (MRI) do joelho, diferenciando imagens normais e anormais associadas à osteoartrite. Utilizando o dataset OAI-MRI-3DDESS, o modelo foi treinado com validação cruzada e oversampling para compensar o desbalanceamento de classes. A abordagem visa aprimorar a precisão diagnóstica e reduzir a sobrecarga dos profissionais, explorando as informações volumétricas das imagens. Os experimentos indicaram desempenho robusto, com acurácia superior a 86% e AUC próximo a 0,92, mostrando o potencial do método para aplicações clínicas.*

1. Introdução

Este estudo aborda o problema da classificação automática de imagens de Ressonância Magnética (MRI) do joelho utilizando aprendizado profundo. O objetivo foi desenvolver um modelo baseado em redes neurais convolucionais 3D (3D CNN) para diferenciar entre imagens normais e anormais utilizando a arquitetura R3D-18 [Hara et al. 2018], aprimorada com aumento de dados dinâmico e transferência de aprendizado para extração de características em volumes tridimensionais, com a base de dados OAI-MRI-3DDESS. A proposta visou reduzir a carga de trabalho dos especialistas, aumentando a precisão diagnóstica, e proporcionar uma ferramenta auxiliar para a detecção precoce de anormalidades. Ela foi motivada pela escassez de métodos generalizáveis para MRI 3D de joelho, onde variações anatômicas e limitações de dados desafiam modelos convencionais [Siouras et al. 2022].

Neste contexto, foi adotada uma abordagem que adapta a arquitetura R3D_18 (3D ResNet-18), originalmente concebida para o reconhecimento de ações em vídeos, para capturar de forma mais precisa as características volumétricas específicas das imagens de MRI. As principais contribuições são: (i) a aplicação de técnicas de *oversampling* para mitigar o desbalanceamento de classes; (ii) o uso de validação cruzada (*k-fold*) para produzir um modelo mais preciso, um treinamento menos sucetível a *overfitting* e melhor aproveitamento dos dados; e (iii) a obtenção de resultados com desempenho discretamente superior aos de outros estudos que empregaram a mesma técnica de adaptação da última camada nesta arquitetura R3D_18. Os resultados indicam o potencial da abordagem para aprimorar o diagnóstico e auxiliar na tomada de decisão clínica.

2. Breve descrição da base de dados

A base de dados OAI-MRI-3DDESS, disponível no Kaggle [Berrimi 2021], contém imagens tridimensionais de ressonância magnética (3D DESS, do inglês, *Three-Dimensional Double-Echo Steady-State*) da articulação do joelho, coletadas de aproximadamente 3.000 pacientes. Essas imagens foram classificadas em positivas ou negativas de acordo com os graus de *Kellgren-Lawrence* (KL), extraídos das radiografias do conjunto de dados complementar da OAI (*Osteoartrites Initiative*).

3. Trabalhos relacionados

Alguns estudos recentes, como [Guida et al. 2021] e [Zhong et al. 2022] têm explorado o uso de técnicas de aprendizado profundo para a classificação de imagens médicas. Por exemplo, pesquisas que utilizaram arquiteturas 2D e 3D mostraram que a adaptação de modelos pré-treinados pode melhorar a extração de características volumétricas, aumentando a acurácia diagnóstica.

Além disso, o estudo de [Al Turkestani et al. 2024] apresenta uma abordagem inovadora para prever a progressão da osteoartrite da articulação temporomandibular (TMJ OA) com base na combinação de biomarcadores clínicos, quantitativos de imagem e biológicos. O modelo desenvolvido, denominado *Ensemble via Hierarchical Predictions through Nested cross-validation* (EHPN), apresentou desempenho superior a 48 modelos testados, atingindo um *F1-score* de 0,82 e uma AUC-ROC (Área Sob a Curva da Característica de Operação do Receptor) de 0,72. Esses achados reforçam a importância da integração de múltiplos tipos de dados na construção de modelos preditivos mais precisos, estratégia que também é explorada neste trabalho ao utilizar uma CNN 3D adaptada para imagens volumétricas de MRI.

Este trabalho se diferencia dos trabalhos encontrados na literatura pela utilização de técnicas de *oversampling* e validação cruzada, *k-fold*, para mitigar o desbalanceamento de classes e garantir a robustez dos modelos. Além disso, pela adaptação da última camada da arquitetura R3D_18 para a classificação binária de imagens de MRI do joelho.

4. Modelo para diferenciação de imagens de joelho

Este modelo utiliza rótulos binários (0 e 1) para classificar imagens de joelho como normais (0) ou anormais (1), empregando uma arquitetura de aprendizado de máquina que será detalhada a seguir. A base de dados disponibiliza apenas rótulos binários relativos à classificação global de cada volume, sem informações de localização (*bounding boxes*) ou segmentação (máscaras de interesse). Logo, métodos de detecção de

objetos, como o YOLO [Redmon et al. 2016], e abordagens de segmentação, como a U-Net [Ronneberger et al. 2015], não são adequados para essa tarefa. Em vez disso, a classificação de cada imagem volumétrica será abordada de forma mais eficaz por meio de CNNs, que extraem automaticamente características discriminativas do volume completo. As imagens são rotuladas como 0 (normal) ou 1 (anormal). A saída y do modelo é convertida em uma probabilidade p, entre 0 e 1, usando a função sigmoide $\sigma(\cdot)$ descrita na Equação (1):

$$p = \sigma(y) = \frac{1}{1 + e^{-y}}. \tag{1}$$

Essa função sigmoide é amplamente utilizada para transformar valores contínuos em probabilidades, em problemas de classificação binária [Goodfellow et al. 2016]. A probabilidade resultante representa a confiança do modelo na classificação da imagem como anormal. Um limiar de 0,5 é então usado para tomar a decisão final da classe, ou seja, se $p \geq 0,5$ então será atribuida a classe **anormal** (1); caso contrário (se $p < 0,5$) será atribuida a classe **normal** (0).

A divisão do conjunto de dados para treinamento, validação e teste correspondem a 70%, 15% e 15% do total de imagens, respectivamente. Isso resulta para a classe normal (0) em 1229, 263 e 267 imagens, e para anormal (1) em 853, 182 e 182 imagens, totalizando 2082, 445 e 449 imagens, respectivamente, para treino, validação e teste.

5. Adaptação do modelo R3D_18

O artigo sobre a detecção do Alzheimer por meio de MRI estrutural [Hosseini-Asl et al. 2016] mostra como uma rede 3D pode ser adaptada para lidar com dados volumétricos de imagens deste contexto. Essa abordagem é diretamente relevante para o presente trabalho, pois evidencia a aplicabilidade de arquiteturas 3D, como a utilizada no R3D_18, para tarefas de classificação de imagens.

Logo, foi adotada a arquitetura R3D_18, originalmente proposta para o reconhecimento de ações em vídeos [Hara et al. 2018], adaptada para a classificação de imagens volumétricas de MRI. Como o R3D_18 está pré-treinado em dados RGB (três canais), foi realizada uma adaptação simples, mas eficaz, para possibilitar o uso original do modelo. No intuito de atender ao requisito dos 3 canais de entrada, foi replicado o único canal de entrada existente, que corresponde às imagens em escala de cinza, para gerar uma entrada de três canais, mantendo a compatibilidade com a estrutura de entrada original do modelo. Além disso, a última camada da rede foi substituída por uma camada totalmente conectada com um neurônio, configurada para a classificação binária entre volumes normais e anormais. Essa estratégia tenta aproveitar a capacidade de extração de características espaciais e temporais do modelo, originalmente destinado a vídeos.

6. *Pipeline* de treinamento com *k-fold*

Em cada *fold*, o conjunto de dados foi dividido em subconjuntos de treinamento e validação, com *oversampling* aplicado aos índices de treinamento para mitigar o desbalanceamento entre as classes, através da replicação aleatória dos elementos da classe numericamente minoritária. Durante o treinamento, foi utilizado um mecanismo de *early stopping* com paciência de 40 épocas, interrompendo o processo de treinamento tão logo as tendências dos desempenhos dos dados de validação e treino apresentem diferença. Cada *fold* teve um tempo médio de execução de 5 a 6 horas, e a implementação também

fez uso de paralelização para aproveitar os recursos de múltiplas GPUs, contribuindo para a eficiência do treinamento [Kohavi 1995]. Para avaliar a consistência do modelo, cada *fold* foi executado três vezes com configurações de inicialização distintas. Os resultados apresentados correspondem à média das execuções. Embora não tenham sido realizados testes estatísticos formais entre os *folds*, as pequenas variações observadas nas métricas sugerem estabilidade. Trabalhos futuros poderão incluir análises de significância (ex: Teste T pareado ou ANOVA) para validar diferenças entre as execuções.

A utilização de *folds* em validação cruzada é dada para obter uma avaliação robusta e generalizável do modelo. Em vez de depender de uma única divisão entre treinamento e validação — que pode introduzir vieses e não refletir a variabilidade dos dados — cada *fold* garante que diferentes subconjuntos da base de dados sejam usados para treinamento e validação. Dessa forma, cada imagem de validação é apreciada e contribui para a avaliação final, reduzindo o risco de *overfitting* e proporcionando uma estimativa mais confiável do desempenho real do modelo. No caso deste trabalho, a escolha de 5 *folds* equilibra o número de experimentos e o volume de dados com o custo computacional disponível, permitindo que o treinamento seja realizado de forma viável. Na Tabela 1 são apresentados os resultados nos 5 *folds* em termos de perda, acurácia, *F1-Score* e AUC-ROC para os dados de validação.

Tabela 1. Resultados dos 5 *folds* do modelo R3D_18 com *early stopping* para o conjunto de dados de validação.

Fold	Perda	Acurácia	F1-Score	AUC-ROC
1	0,111	76,0%	0,731	0,852
2	0,097	83,4%	0,817	0,904
3	0,121	82,5%	0,804	0,889
4	0,093	82,9%	0,804	0,900
5	0,098	78,0%	0,774	0,874

O *fold* 2 apresentou o melhor desempenho geral na classificação das imagens da base de dados. Ele obteve a maior acurácia (83,36%), o maior *F1-Score* (0,8170) e o maior valor AUC-ROC (0,9036), indicando alta precisão e equilíbrio entre as classes. O *fold* 4 apresentou a menor perda de teste (0,0932), mas teve desempenho inferior nas demais métricas. Dessa forma, ao considerar todas as métricas de avaliação, o *fold* 2 é o mais eficiente na diferenciação entre imagens normais e anormais, demonstrando que o modelo aprendeu padrões relevantes das imagens de ressonância magnética do joelho. Caso a prioridade seja apenas a minimização da perda, o *fold* 4 poderia ser considerado, mas, para um desempenho globalmente superior, o *fold* 2 seria a melhor escolha.

7. Último modelo utilizando o conjunto inteiro de treino

Após a realização de validação cruzada em 5 *folds*, identificou-se que o *fold* 2 apresentou o melhor desempenho com base nas métricas de validação. Os hiperparâmetros são configurações que controlam o treinamento do modelo. A *learning rate* (0,0001) ajusta a velocidade de aprendizado, evitando ajustes excessivos. O *batch size* (8) define quantos dados de entrada são processados por vez, ajudando a evitar *overfitting*. O *weight decay* (10^{-5}) regulariza o modelo, penalizando pesos grandes. A *focal loss* (`alpha`=0,25, `gamma`=2, e `pos_weight`=2) lidam com desbalanceamento de classes, focando em exemplos difíceis. O *oversampling* balanceia as classes, aumentando amostras da classe minoritária. Esses ajustes melhoram o desempenho e a generalização do modelo. Neste

treinamento, o critério para definir a melhor época foi baseado no desempenho do *fold* 2, que teve o melhor resultado durante a validação cruzada. O *fold* 2 atingiu seu melhor resultado na época 53, com uma perda (*loss*) de validação de 0,097, acurácia de 83,4%, AUC-ROC de 0,904 e *F1-Score* de 0,817. Desta forma, o modelo final foi treinado usando as mesmas 53 épocas e todos os hiperparâmetros selecionados pelo desempenho do *fold* 2. Após o treinamento no conjunto completo (treino + validação), o modelo alcançou uma acurácia de 86,1%, AUC-ROC de 0,920 e *F1-Score* de 0,834 quando aplicado ao conjunto de teste, representando uma melhoria significativa em relação aos resultados originais do *fold* 2.

8. Comparação dos resultados

O modelo 3D CNN utilizado no estudo de [Guida et al. 2021] foi inspirado na arquitetura ResNet-50, adaptada para o processamento de imagens de ressonância magnética 3D. Esse modelo foi desenvolvido para analisar sequências de imagens de MRI, aproveitando a capacidade dos convolucionais 3D para extrair características volumétricas, ou seja, informações contextuais extraídas de cortes adjacentes das imagens. Isso permitiu a captura de características que não seriam detectadas por modelos convencionais 2D CNN. Já no artigo de [Zhong et al. 2022], os autores empregaram conjuntos de dados de MRI de *Osteoarthritis Initiative* (OAI) e utilizaram a pontuação MRI do *Osteoarthritis Knee Score* (MOAKS) para rotular áreas de cartilagem, como a patelar, femoral, medial tibial e lateral tibial. A descrição do método foi composto por duas etapas principais, a primeira envolvendo a segmentação das regiões de interesse por meio de uma U-Net treinada com dados 3D DESS, e a segunda etapa utilizando variantes dos modelos DenseNet (DenseNet 121 e DenseNet 169). Com isso, a Tabela 2 apresenta uma comparação entre os resultados obtidos neste estudo e as abordagens propostas por esses outros trabalhos, que utilizaram o mesmo tipo de base de dados referenciado anteriormente.

Tabela 2. Comparação de resultados entre os artigos analisados e o presente trabalho, onde os campos "Referência": cita o trabalho associado aos resultados, "Dataset": todos com dados de OIA-MRI-3DDESS e diferentes números de imagens, "Teste": número de imagens de teste, "Classes": todos binários com diferenças na rotulagem com OA igual a Normal baseado nos graus KL e MOAKS, "Balanceamento": diferença entre classes equilibradas, desbalanceadas e artificialmente equilibradas. Além das métrica "Acurácia", "AUC-ROC" e "F1-Score".

Referência	[Guida et al. 2021]	[Zhong et al. 2022]	Trabalho proposto
Dataset	1.100 imagens	2.396 imagens	2.976 imagens
Teste	100 imagens	480 imagens	449 imagens
Modelo	3D ResNet-50	3D DenseNet 121 e 169	3D ResNet-18
Classes	OA×não-OA	MOAKS×não-MOAKS	Normal×Anormal
Balanceamento	Balanceada	Desbalanceada	*Oversampling*
Acurácia	83,0%	75% - 83%	86,1%
AUC-ROC	0,911	- - - [1]	0,920
F1-Score	0,831	0,71 - 0,89 [2]	0,834

Os resultados deste trabalho indicam um desempenho competitivo em relação às abordagens anteriores, com uma acurácia de 86,1% no modelo treinado com o conjunto de treinamento completo, comparável aos melhores resultados obtidos nos artigos analisados. Além disso, as métricas de AUC-ROC e *F1-Score* foram 0,920 e 0,834, respectivamente. O modelo 3D CNN adaptado, baseado na arquitetura R3D_18, atingiu o

[1] Valores não fornecidos.

[2] Valores estimados a partir dos dados fornecidos.

objetivo de aprimorar a classificação binária de imagens de MRI, mostrando resultados promissores. Esses achados indicam que o modelo pode ser uma ferramenta eficaz para a detecção de anomalias em imagens de MRI, com o potencial de auxiliar no diagnóstico em ambientes clínicos.

9. Conclusão

Este estudo verificou a viabilidade de uma rede neural convolucional 3D (*R3D_18*) para classificação binária de imagens de MRI do joelho, com *86,1% de acurácia* e *AUC-ROC de 0,920*, sugerindo aplicabilidade na detecção precoce de osteoartrite. A extração de informações volumétricas, combinada a técnicas de *oversampling* e validação cruzada (*5-folds*), assegurou robustez ao modelo, mesmo com desbalanceamento de classes. Como limitações, destacam-se a necessidade de ampliação do *dataset* e comparação com métodos alternativos. Futuramente, a integração de biomarcadores clínicos poderá aprimorar a predição, direcionando estratégias terapêuticas personalizadas. A abordagem proposta reduziria a demanda por análises manuais e fundamenta avanços em diagnósticos automatizados de OAI.

Referências

Al Turkestani, N., Li, T., Bianchi, J., Gurgel, M., Prieto, J., Shah, H., Benavides, E., Soki, F., Mishina, Y., and Fontana, M. (2024). A comprehensive patient-specific prediction model for temporomandibular joint osteoarthritis progression. *Proc. Natl. Acad. Sci. (PNAS)*, 121(8):e2306132121.

Berrimi, M. (2021). OAI-MRI-3DDESS Dataset. Acesso em: nov. 2024. Disponível em: `https://www.kaggle.com/datasets/mohamedberrimi/oaimri3ddess`. Dataset público de ressonância magnética 3D do joelho.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Convolutional networks*. MIT press Cambridge.

Guida, C., Zhang, M., and Shan, J. (2021). Knee osteoarthritis classification using 3D CNN and MRI. *Applied Sciences*, 11(11):5196.

Hara, K., Kataoka, H., and Satoh, Y. (2018). Can spatiotemporal 3D CNNS retrace the history of 2D CNNs and ImageNet? In *Proc. 2018 IEEE/CVF Conf. CVPR*, pages 6546–6555, Salt Lake City, USA.

Hosseini-Asl, E., Keynton, R., and El-Baz, A. (2016). Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In *Proc. 2016 IEEE Intl. Conf. Img. Process. (ICIP)*, pages 126–130, Phoenix, USA.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *14th Intl. Joint Conf. AI - IJCAI 1995*.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once. In *Proc. 2016 IEEE/CVF Conf. CVPR*, pages 779–788, Las Vegas, USA.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Proc. 18th Intl. Conf. Med. Img. Comp. & Comp.-Asst. Interv. – MICCAI 2015*, pages 234–241, Munich, Germany.

Siouras, A., Moustakidis, S., Giannakidis, A., Chalatsis, G., Liampas, I., Vlychou, M., Hantes, M., Tasoulis, S., and Tsaopoulos, D. (2022). Knee injury detection using deep learning on mri studies: a systematic review. *Diagnostics*, 12(2):537.

Zhong, J., Yao, Y., Khan, S., Xiao, F., Cahill, D. G., Griffith, J. F., and Chen, W. (2022). Knee Osteoarthritis: Automatic Grading with Deep Learning. In *Proc. 2022 ISMRM & ISMRT Annu. Mtg. Expo.*