UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA
Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto n°17.228 de 25/11/2016
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXVII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS
SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2023

ANÁLISE DE INTERPRETABILIDADE PARA OS CLASSIFICADORES DO SISTEMA PATHOSPOTTER

Silas Silva da Costa[1] e Angelo Amancio Duarte[2]

1. Bolsista PIBIC/CNPq, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: silllas22@gmail.com

2. Orientador, Departamento de Tecnologia, Universidade Estadual de Feira de Santana, e-mail: angeloduarte@uefs.br

PALAVRAS-CHAVE: Computional Vision; Explainable Intelligence; Computational Pathology.

INTRODUÇÃO

O campo da patologia computacional se concentra na aplicação da computação para melhorar a análise e interpretação de dados em patologia. Isso envolve a utilização de recursos, como machine learning e visão computacional. Um método amplamente utilizado nesse contexto é o emprego de Convolutional Neural Networks (CNN), ou Redes Neurais Convolucionais em português, especialmente em tarefas relacionadas à análise de imagens.

Um exemplo notável de aplicação de CNN é o PathoSpotter-Search, desenvolvido por Aguiar et al. (2021), que é um classificador de lesões renais glomerulares. No entanto, existe impedimentos na adoção de técnicas como CNN (MAIA, 2022), Isso se deve ao fato de que não é possível compreender completamente os critérios e as características específicas que são usados para identificar e classificar imagens por meio de CNNs. Essas redes neurais são frequentemente descritas como "caixas pretas" porque, dado um conjunto de entradas, geram saídas sem fornecer uma explicação do sobre o que ocorreu internamente (GUNNING; AHA, 2019).

Diante desse problema, o presente trabalho propõe a utilização de técnicas de interpretabilidade de CNNs. O objetivo é identificar quais critérios ou partes específicas da imagem foram utilizados no processo de tomada de decisão, tornando as operações das redes neurais mais transparentes e compreensíveis para os profissionais de saúde e pesquisadores.

METODOLOGIA

Com o objetivo de tornar as operações das Redes Neurais Convolucionais (CNN) mais transparentes e compreensíveis para profissionais de saúde e pesquisadores no contexto da classificação de imagens de glomérulos renais, foram selecionadas quatro técnicas de interpretabilidade de CNNs. Essas técnicas permitem identificar quais critérios ou partes específicas das imagens foram utilizados no processo de tomada de decisão da CNN.

Essas técnicas oferecem visualização das partes mais importantes da imagem, em outras palavras das partes decisivas para que a CNN classifique a imagem em uma determinada classe. Cada técina tem sua própria maneira de execução, o que fica claro quando o resultado de cada uma é plotado lado a lado.

1. GradCAM (Gradient-weighted Class Activation Mapping)

Esta técnica foi proposta por SELVARAJU et al. (2017), e visa destacar as regiões de uma imagem que mais influenciaram a decisão da CNN, oferecendo uma interpretação visual das áreas de importância.

2. GradCAM++ (GradCAM Plus Plus)

Proposta por CHATTOPADHYAY et al. (2017), é uma extensão do GradCAM, esta técnica aprimora a localização das regiões relevantes, fornecendo uma visão mas focada das características importantes.

3. SmoothGrad

Proposta por SMILKOV et al. (2017) essa técnica aplica um processo de suavização nas entradas da rede neural, adicionando ruído aleatório a uma imagem de entrada varias vezes e calculando as medias das saídas da rede em relação a essas imagens perturbadas.

4. Mapa de Saliência (Saliency Map)

Proposta por SIMONYAN; VEDALDI & ZISSERMAN (2014), utilizando principios de atenção visual, esta técnica identifica as áreas de uma imagem que atraem a atenção da CNN, ajudando a compreender quais características são mais relevantes para a classificação.

Cada uma das técnicas mencionadas foi aplicada à CNN treinada para classificação de imagens de glomérulos renais, com o intuito de fornecer uma análise das características determinantes na classificação. Essas técnicas de interpretabilidade contribuíram para tornar o processo de decisão da CNN mais transparente, possibilitando uma melhor compreensão de quais as áreas da imagem eram mais relevantes para a decisão do modelo.

RESULTADOS E/OU DISCUSSÃO

Foram implementados quatro métodos de visualização que se revelaram eficazes na geração de representações visuais dos resultados das Redes Neurais Convolucionais. Para exemplificar a capacidade dos métodos, consideramos uma imagem (Figura 1) contendo uma imagem com lesão cuja previsão foi corretamente feita pelas CNNs. Os resultados das visualizações estão apresentados como imagens sobrepostas com mapa de calor, conforme ilustrado na Figura 2, fazendo referencia ao resultado de uma CNN e na figura 3, correspondente a outra CNN, para fins de identificação, as CNNs foram nomeadas de CNN 1 e CNN 2, respectivamente.

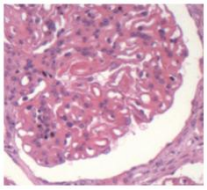

Figura 1: Glomérulo com lesão

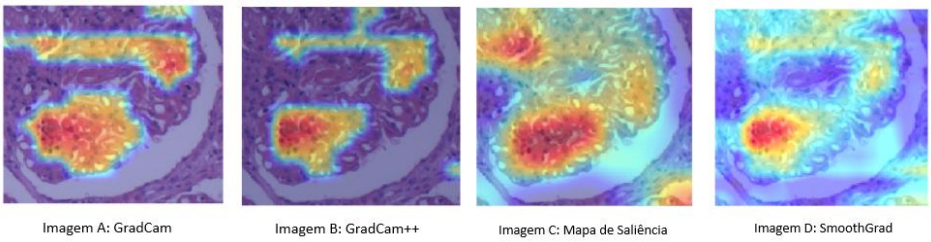

Figura 2: Resultado das técnicas aplicadas na CNN 1

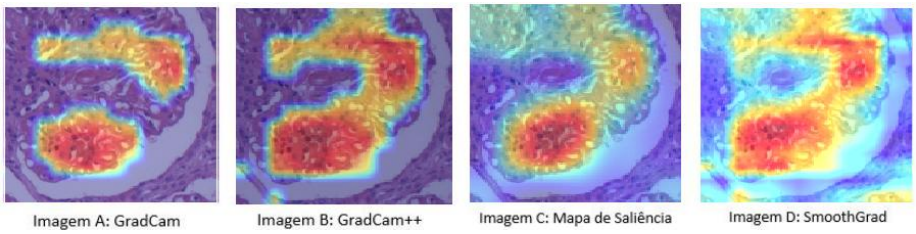

Figura 3: Resultado das técnicas aplicadas na CNN 2

Resultado nas imagens abaixo (Figura 5) e (Figura 6) das técnicas aplicados nas CNNs na classificação da imagem (Figura 4) uma imagem sem lesão classificada corretamente pela CNN.

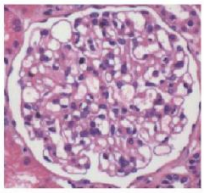

Figura 4: Glomérulo sem lesão

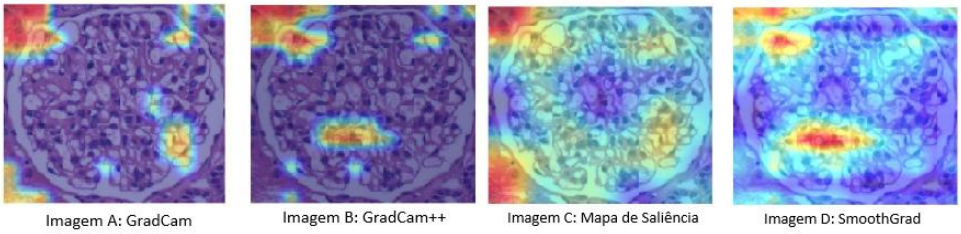

Figura 5: Resultado das técnicas aplicadas na CNN 1

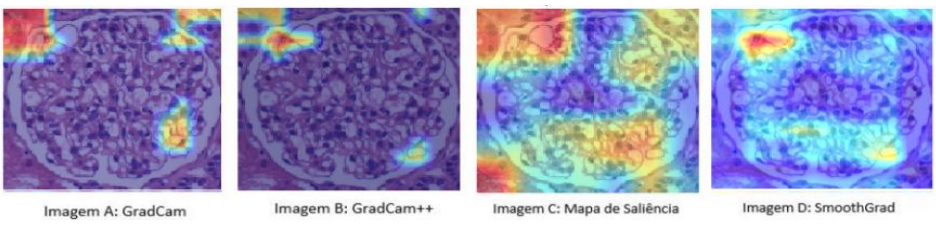

Figura 6: Resultado das técnicas aplicadas na CNN 2

CONSIDERAÇÕES FINAIS (ou Conclusão)

Neste estudo, foi explorada a implementação de quatro métodos de visualização para analisar e interpretar os resultados obtidos por Redes Neurais Convolucionais (CNNs). Foi demonstrado que esses métodos são capazes de gerar representações visuais informativas das previsões das CNNs.

As visualizações resultantes, apresentadas na forma de imagens sobrepostas com mapas de calor, proporcionaram uma compreensão de quais partes da imagem mais interfere no processo de tomada de decisão das CNNs.

Mesmo com as dificuldades encontradas durante o processo de desenvolvimento, foi possível ter um resultado satisfatório na aplicação de técnicas de visualização. O trabalho ainda tem espaço para melhorias, como otimização das técnicas descritas, implementação de novas técnicas de visualização.

A oportunidade oferecida pelo projeto de iniciação científica foi de extrema importância para o desenvolvimento de habilidades relacionadas a Machine Learning, Redes Neurais Convolucionais e a interpretabilidade dessas redes. Isso contribuiu significativamente para a aquisição e aplicação de conhecimentos específicos.

REFERÊNCIAS

AGUIAR, E. et al. PathoSpotter-Search: A Content-Based Image Retrieval Tool for Nephropathology. In: Anais Estendidos do XXXIV Conference on Graphics, Patterns and Images. SBC, 2021. p. 146-149.

CHATTOPADHYAY, Aditya; **SARKAR**, Anirban; **HOWLADER**, Prantik; **BALASUBRAMANIAN**, Vineeth. Grad-CAM++: Generalized Gradient-based Visual Explanations for Deep Convolutional Networks, 2017.

GUNNING, David; **AHA**, David W. DARPA's Explainable Artificial Intelligence Program. AI Magazine, La Canada, v. 40, n. 2, p. 44-58, verão 2019.

MAIA, Matheus Gomes. Interpretabilidade de redes neurais convolucionais com estudo de caso em diagnóstico por imagem. Campina Grande, 2022. 107 f. il. color. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2022. Orientação: Prof. Dr. Herman Martins Gomes.

SELVARAJU, Ramprasaath R. et al. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618-626.

SIMONYAN, Karen; **VEDALDI**, Andrea; **ZISSERMAN**, Andrew. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. ICLR (Workshop Poster), 2014.

SMILKOV, Daniel et al. SmoothGrad: removing noise by adding noise. CoRR, abs/1706.03825, 2017.