UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA
Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº17.228 de 25/11/2016
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXIX SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS
SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2025

Investigação de mecanismos de atenção para a construção de classificadores de imagens de glomérulos renais

Lyrton Marcell Dias Amorim[1]; Angelo Amancio Duarte[2]
1.Lyrton Marcell Dias Amorim, PROBIC/UEFS, ENGENHARIA DE COMPUTAÇÃO, e-mail: lymarcell@gmail.com
2.Angelo Amancio Duarte, DEPARTAMENTO DE TECNOLOGIA, e-mail: angeloduarte@uefs.br

PALAVRAS-CHAVE: Mecanismo de Atenção; Aprendizado de Máquina; Classificação de Imagens Médicas.

INTRODUÇÃO

O avanço do aprendizado profundo tem permitido a criação de modelos capazes de auxiliar no diagnóstico de doenças renais a partir de imagens histológicas, destacando-se os mecanismos de atenção, que direcionam o foco do modelo para regiões mais relevantes da imagem. No caso das lesões glomerulares, como hipercelularidade, esclerose e amiloidose, esse recurso se mostra promissor. Nesse contexto, arquiteturas modernas como o Vision Transformer (ViT) e suas variações, que exploram relações espaciais e contextuais, apresentam grande potencial, mas também demandam grandes quantidades de dados rotulados, o que é um desafio em aplicações médicas. Diante disso, este trabalho tem como objetivo investigar o uso de mecanismos de atenção na construção de classificadores de lesões glomerulares em biópsias renais digitais, buscando compreender seu funcionamento, implementar modelos baseados nessa abordagem e comparar seus resultados com soluções já desenvolvidas, como as do projeto PathoSpotter, de modo a avaliar as contribuições e limitações dessa tecnologia no apoio ao diagnóstico.

METODOLOGIA

Figura 1: Arquitetura do *Multi-Path Vision Transformer*. Fonte: Keras.

A pesquisa foi desenvolvida a partir de experimentos computacionais aplicados à classificação de lesões glomerulares em imagens digitais de biópsias renais. O conjunto de dados utilizado apresentou forte desbalanceamento entre as classes, refletindo a realidade clínica em que algumas lesões são raras. As imagens foram redimensionadas

para 224 × 224 pixels, normalizadas e padronizadas (Almeida, 2023), garantindo compatibilidade com a arquitetura utilizada, para mitigar o desbalanceamento, aplicaram-se técnicas de aumento de dados (data augmentation), undersampling e oversampling, gerando maior variabilidade e equilíbrio entre as classes.

Como modelo de classificação, implementou-se o Global Context Vision Transformer (GCViT) (Hatamizadeh et al., 2022), uma arquitetura híbrida que combina convoluções locais e mecanismos de atenção global, capaz de capturar padrões sutis e contextos mais amplos das imagens. O treinamento foi realizado com validação cruzada em 5 folds (Berrar, 2019), possibilitando avaliar a robustez do modelo e reduzir o risco de sobreajuste. Esse procedimento metodológico permitiu analisar de forma sistemática o desempenho da arquitetura escolhida em diferentes condições do dataset.

RESULTADOS E/OU DISCUSSÃO

Nesta seção, são apresentados e discutidos os resultados obtidos nos experimentos realizados com a arquitetura Vision Transformer com Contexto Global (GCViT). Os testes foram conduzidos em três cenários distintos: dataset desbalanceado, dataset balanceado por undersampling e dataset balanceado por oversampling com data augmentation. O objetivo foi avaliar como diferentes estratégias de balanceamento impactam a capacidade do modelo de generalizar e identificar padrões relevantes em imagens histológicas de lesões glomerulares.

A. Dataset desbalanceado com Vision Transformer Contexto Global

No primeiro experimento, o modelo foi treinado com o dataset completo e desbalanceado. A Figura 1 mostra a evolução do erro de treino ao longo das épocas nos cinco folds de validação cruzada. Observa-se que todos os folds apresentaram queda consistente do erro, indicando aprendizado durante o processo. No entanto, o F1-score final foi de apenas 6%, revelando que, mesmo com baixa perda no treinamento, o modelo não conseguiu generalizar para dados novos. Esse comportamento está associado ao sobreajuste (overfitting), consequência direta do forte desbalanceamento das classes. Assim, o modelo "memoriza" padrões da classe majoritária e falha em classificar corretamente as minoritárias.

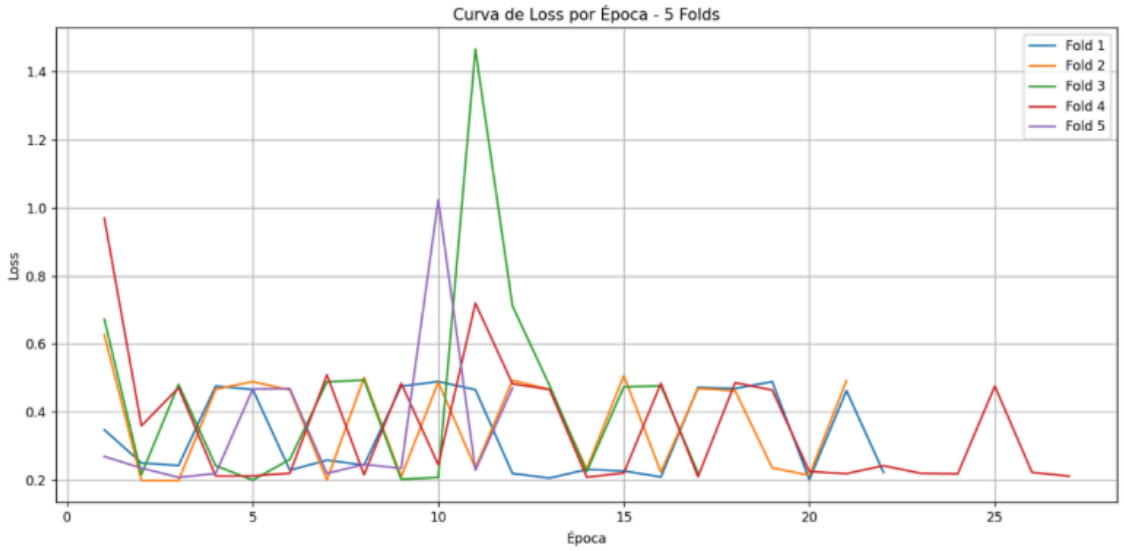

Figura 2: Evolução do erro de treino nos 5 folds utilizando o dataset desbalanceado.

B. Dataset com undersampling e Vision Transformer Contexto Global

No segundo experimento, aplicou-se a técnica de undersampling para reduzir o desbalanceamento. A Figura 2 apresenta a evolução do erro de treino nos cinco folds. Nota-se que as curvas se mostram mais estáveis e com menos oscilações em comparação ao primeiro cenário. Esse comportamento reflete um aprendizado mais consistente, já que o modelo não foi tão influenciado pela classe majoritária. O desempenho geral melhorou, alcançando F1-score de 49,29%, o que representa um avanço significativo em relação ao dataset desbalanceado. Apesar disso, o uso do undersampling reduziu a quantidade total de amostras de treino, o que limitou o potencial máximo de aprendizado do modelo.

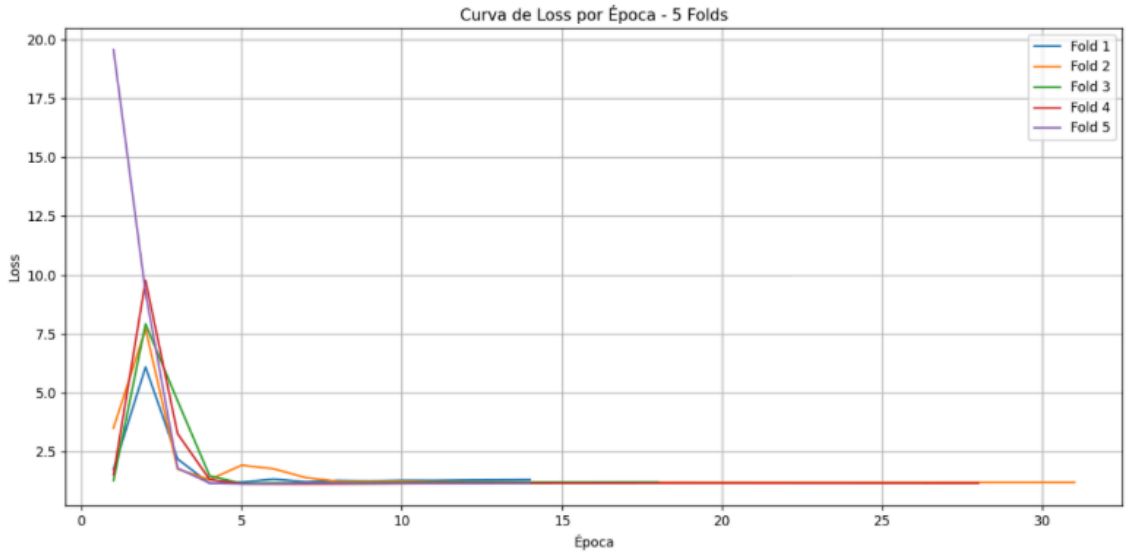

Figura 3: Evolução do erro de treino nos 5 folds utilizando undersampling.

C. Dataset com oversampling + data augmentation e Vision Transformer Contexto Global

No terceiro experimento, utilizou-se oversampling aliado ao data augmentation para aumentar a representatividade das classes minoritárias. A Figura 3 mostra que as curvas de erro convergiram de forma estável e sem grandes oscilações, chegando a valores muito próximos de zero. Esse resultado evidencia que o modelo conseguiu aproveitar a maior variedade de exemplos disponíveis para aprender de maneira mais robusta. O F1-score final de 67,73% foi o melhor desempenho entre os cenários avaliados, mostrando que o oversampling, quando combinado com data augmentation, permitiu ao modelo capturar padrões mais gerais e reduzir o viés em relação às classes minoritárias.

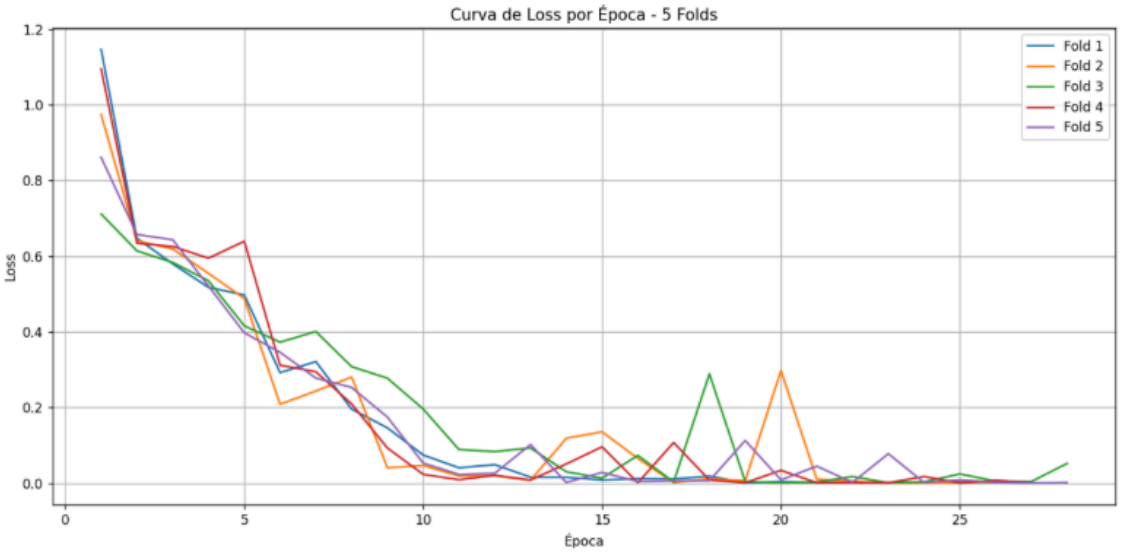

Figura 4: Evolução do erro de treino nos 5 folds utilizando oversampling e data augmentation.

CONSIDERAÇÕES FINAIS

Este trabalho analisou o uso de modelos de atenção na classificação de lesões glomerulares em imagens digitais de biópsias renais, com foco no Vision Transformer com Contexto Global. Os resultados mostraram que, embora a arquitetura apresente potencial para capturar padrões visuais complexos, seu desempenho foi limitado pelas condições do estudo, especialmente pelo desbalanceamento das classes e pela escassez de dados. Além disso, a alta demanda computacional e a sensibilidade a hiperparâmetros reforçam os desafios do uso desse modelo em cenários médicos reais. Ainda assim, os experimentos indicam que estratégias como ampliação do dataset, pré-treinamento em domínios relacionados, aprendizado auto-supervisionado e métodos de interpretabilidade podem ser caminhos promissores para tornar essas arquiteturas mais eficazes e aplicáveis na prática clínica.

REFERÊNCIAS

ALMEIDA, R. 2023. Padronização de dados em redes neurais convolucionais aplicadas à classificação de imagens médicas. Universidade Estadual de Feira de Santana, Relatório Técnico.

BERRAR, D. 2019. Cross-Validation. In: RENSKI, E. (ed.) Encyclopedia of Bioinformatics and Computational Biology. Oxford, Elsevier, p. 542–545.

HATAMIZADEH, A.; KOSIŃSKI, W.; DOU, Q.; ERMON, S.; LUO, J. 2022. Global Context Vision Transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), p. 14984–14994.

LIN, A.; GAN, Z.; HANEY, M.; LIU, Z.; WANG, L.; LI, J. 2022. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), p. 10012–10022.

ZHOU, T.; FU, H.; HAN, X.; DONG, B.; ZHANG, Y. 2023. Global–Local Transformer for Medical Image Segmentation. Medical Image Analysis, 87: 102840.