

Aprendizado Profundo na Classificação de Lesões Crescentes Glomerulares: modelos e condições

Joacy Mesquita da Silva*, Michele Fulvia Angelo*, Washington L. C. dos Santos[†], e Angelo C. Loula*

*Universidade Estadual de Feira de Santana (UEFS), Bahia, Brasil

Email: joacymasilva@gmail.com, mfangelo@ecom.uefs.br, angelocl@uefs.br

[†]Fundação Oswaldo Cruz - Instituto Gonçalo Moniz, Bahia, Brasil

Email: wluis@bahia.fiocruz.br

Abstract—Glomeruli are structures in the kidneys, responsible for filtering the blood, that can be affected by several lesions, such as the glomerular crescent, which is characterized by abnormal cell proliferation. In this work, different models and conditions for the application of deep learning are to be evaluated in the task of classifying glomerular crescent histopathological images. The pre-trained networks Xception, InceptionV3, MobileNet, VGG16 and ResNet50 were compared, by applying to the classification of images with crescent vs normal glomeruli. Comparing the accuracy, precision, recall and f1-score of the models, the ResNet50 showed significantly better performance than the other networks, in all measures. The application of data augmentation did not result in a significant improvement in the results in this case. In an experiment of classification of crescent vs non-crescent glomeruli, adding images of three other lesions to the database, the application of Focal Loss presented greater accuracy and precision.

Resumo—Glomérulos são estruturas localizadas nos rins e responsáveis por filtrar o sangue e podem ser acometidos por diversas lesões, como a crescente glomerular, que é caracterizada por apresentar uma anormal proliferação de células. Neste trabalho, são avaliados diferentes modelos e condições de aplicação de deep learning na tarefa de classificação de imagens histopatológicas de crescente glomerular. Para isso, foram comparadas as redes pré-treinadas Xception, InceptionV3, MobileNet, VGG16 e ResNet50, aplicando-se para a classificação de imagens com glomérulos com crescente vs normais. Comparando acurácia, precisão, recall e f1-score dos modelos, a ResNet50 apresentou desempenho significativamente superior ao das demais redes, em todas as medidas. A aplicação de *data augmentation* não resultou em melhora significativa nos resultados neste caso. Em experimento de classificação de glomérulos crescentes vs não crescentes, adicionando imagens de três outras lesões à base de dados, a aplicação do Focal Loss, apresentou maior acurácia e precisão.

I. INTRODUÇÃO

Na patologia renal ou nefropatologia, os glomérulos, estruturas localizadas no córtex renal e que são responsáveis pela realização da filtração sanguínea, são as regiões em que normalmente se concentram os indicadores de doença, estando sujeitos a padrões diversos de lesões. Assim uma etapa importante do diagnóstico de inúmeras patologias renais é avaliar lesões glomerulares em lâminas histológicas [1]. Dentre as variadas lesões capazes de afetar os glomérulos, há a hiperplasticidade, com aumento na quantidade de células nos glomérulos [1]; a glomeruloesclerose, com alterações morfológicas dos glomérulos pelo endurecimento patológico dos tecidos [2], as lesões membranosas, com o espessamento da

parede dos capilares glomerulares [3], e a crescente glomerular, caracterizada por lesões que apresentam uma anormal proliferação de células, envolvendo 10% ou mais da circunferência da cápsula de Bowman. A presença de crescentes em biópsias renais pode acontecer em diversos distúrbios glomerulares, denominados glomerulopatias com crescentes [4].

Vários trabalhos de aplicações de *machine learning* e *deep learning* na patologia digital por análise de imagens da nefrologia estão presentes na literatura [1], [2], [5]–[10]. As técnicas utilizadas pelos autores dos trabalhos são variadas, foram identificadas pesquisas abordando problemas de classificação, detecção, e segmentação. Barros et. al. [5] utilizam métodos clássicos de processamento de imagem e o *k-nearest neighbors* para a classificação de lesões glomerulares proliferativas. Marsh et. al. [6] apresentam um modelo de aprendizado profundo para identificar e classificar glomeruloesclerose em WSI de biópsias de seção congelada de rim de doadores. Ginley et. al. [7] projetam um pipeline computacional para tarefas de detecção, segmentação, extração de características, e classificação (através de uma rede neural recorrente) da glomeruloesclerose diabética. Chagas et. al. [1] propõem uma arquitetura composta por uma *Convolutional Neural Network* (CNN) e uma *Support Vector Machine* (SVM) para a classificação, binária e de sublesões, de hiperplasticidade glomerular. Bueno et. al. [2] apresentam a segmentação semântica baseada em CNNs para detecção de glomérulos em WSI, seguida por uma CNN para a classificação dos glomérulos em normais ou esclerosados. Uchino et. al. [8] projetam modelos de aprendizado profundo para a classificação de sete lesões distintas: esclerose global, esclerose segmentar, proliferação endocapilar, acúmulo de matriz mesangial, proliferação de células mesangiais, crescente e mudanças estruturais da membrana basal. Chagas et. al. [9] fazem uso de três arquiteturas baseadas em deep learning para a classificação de nefropatia membranosa, em um dataset com imagens de glomérulos normais, de glomérulos com nefropatia membranosa, com hiperplasticidade e com esclerose, e obtiveram resultados de F1-score acima de 92% para os três modelos.

Dentre os trabalhos anteriores, a maioria trata de classificação entre glomérulos normais e glomérulos com determinada lesão. Alguns trabalhos apresentam a classificação normal vs esclerose [2], [6], [7]. Outros concentram-se na

hipercelularidade glomerular como lesão de interesse [1], [5]. Apenas Chagas et. al. [9] foram além de tarefa de distinguir imagens normais de imagens de uma certa lesão, e consideraram as classes de membranosa, outras lesões, e normais. Somente Uchino et. al. [8] abordou a lesão de crescente glomerular e para cada lesão abordada na pesquisa, a classificação é do tipo lesão vs todas as outras lesões. Nenhum dos artigos citados explicitou uma metodologia para tratamento do problema de desbalanceamento das classes, contudo, [8] utilizou ajuste fino em uma rede InceptionV3 para a classificação de dados bastante desbalanceados.

O objetivo deste trabalho é avaliar diferentes modelos e condições de aplicação de *deep learning* na tarefa de classificação de imagens histopatológicas de crescente glomerular. Para isso, são considerados diferentes modelos de redes profundas pré-treinadas aplicando-se *transfer learning*, para a classificação de imagens com glomérulos com crescente vs normais (não-lesionados), e comparando-se também as condições de *data augmentation*. Além desta tarefa, buscando aproximar uma situação mais real da aplicação, foram incluídas imagens com outras lesões glomerulares ao conjunto, e foi avaliada a condição de tarefa imagens com glomérulos com crescente vs não crescente (normais + outras lesões). Em ambas condições, há desbalanceamento de imagens por classe, um tema que ainda demanda estudos em *deep learning* [11], assim são avaliadas alguns possibilidades de tratamento de desbalanceamento por mudanças na função de perda (*loss*).

II. DESCRIÇÃO DO EXPERIMENTO

A. Base de dados

Foi utilizada uma base de dados composta de 6507 imagens histológicas de glomérulos de rins de humanos, das quais 1189 são glomérulos com crescentes, 1571 são glomérulos normais, 1172 são glomérulos com esclerose, 1091 são glomérulos com hipercelularidade, e 1484 são glomérulos com membranosa. As imagens foram fornecidas pelo Centro de Pesquisas Gonçalo Muniz da Fundação Oswaldo Cruz (CpqGM/FIOCRUZ). Elas foram adquiridas através de uma câmera fotográfica digital acoplada a um microscópio. Não houve nenhuma padronização para esta aquisição. A base de dados possui imagens coradas por AZAN, H&E (Hematoxiline & Eosin), PAMS (Periodic Acid Methenamine Silver), PAS (Periodic Acid Schiff), PS (Periodic Schiff), e PICRO. A Figura 1 contém exemplos das imagens presentes na base de dados.

As dimensões das imagens são diversas, as alturas variam de 205 a 3286 pixels, e as larguras variam de 238 e 4128 pixels. Para os experimentos, as imagens foram redimensionadas para um tamanho padrão, a fim de minimizar custos de processamento, as dimensões adotadas foram de 128 x 128 pixels.

B. Tarefa e Modelo de Deep Learning

A proposta do trabalho foi a de avaliar condições de aplicação de *deep learning* na tarefa de classificação de imagens com glomérulos com crescente.

A primeira tarefa de classificação proposta foi de classificação entre imagens de crescente e imagens de glomérulo normal (sem lesões). Foi aplicado *transfer learning*, através do uso de redes pré-treinadas no desafio ImageNet, como forma de extrair características robustas e reduzir o tempo de treinamento do modelo, com um volume menor de dados. Nesta tarefa foram comparadas as redes Xception, InceptionV3, MobileNet, VGG16 e ResNet50. A arquitetura dos modelos utilizados é composta por um modelo pré-treinado, uma camada totalmente conectada, e uma camada de ativação “softmax”. Uma abordagem supervisionada foi utilizada como forma de treinamento desses modelos. Foi realizada também uma avaliação de incremento sintético da base de dados de treinamento pelo uso de técnicas de *data augmentation*, inserindo novas versões das imagens da base de dados, resultantes de transformações geométricas, e também de transformações no espaço de cor. Nos experimentos realizados, as imagens sofreram rotações de 90°, 180°, e 270°, e perda de saturação em 10%, 30%, e 75%, cada imagem originou seis outras imagens. A função de perda utilizada foi entropia cruzada.

Para a segunda tarefa, considerando que em uma situação real diversos tipos de lesões glomerulares além de crescente podem ser encontradas em um aplicação real, a classificação deve distinguir entre imagens de crescente e imagens de não-crescente, que incluem os glomérulos normais, com hipercelularidade, com esclerose, e com lesão membranosa. Este experimento foi realizado com a rede ResNet50, selecionada pelo desempenho na primeira tarefa de classificação. Isto implicou em um forte desbalanceamento de classes, já que 1189 imagens pertencem à classe dos crescentes, e 5318 imagens à classe dos não crescentes. Para isso, duas abordagens de tratamento pelo ajuste da função de perda (*loss function*) foram adotadas: uso de pesos por classe na entropia cruzada ponderada; e o ajuste dinâmico da entropia cruzada por Focal Loss [12], que balanceia e dinamicamente modula a função conforme dificuldade/facilidade de classificação dos exemplos de treino.

Nos dois casos, o framework Keras foi utilizado para modelar, treinar e avaliar os modelos. As redes foram treinadas por 100 épocas, o otimizador Adam foi adotado em todos os experimentos, com taxa de aprendizagem 0.001, o batch size foi de 32. Os experimentos foram realizados na plataforma do Google Colab, com máquina configurada com 25.46 GB de RAM, e GPU NVIDIA-SMI 470.57.02.

A avaliação dos modelos e estimativa de medidas de desempenho nos experimento utilizou validação cruzada com 10 folds (10CV). Os pesos iniciais dos modelos a cada fold foram os mesmos. Os dados coletados através do experimento se relacionam à quatro medidas que serão extraídas com a execução dos modelos: acurácia, precisão, recall, e F1-score. Foi feita a aplicação do teste estatístico de Wilcoxon com intervalo de confiança de 95% para comparação da desempenho dos modelos na tarefa de classificação dos glomérulos, foram obtidas 20 amostras de cada medida (2x10CV).

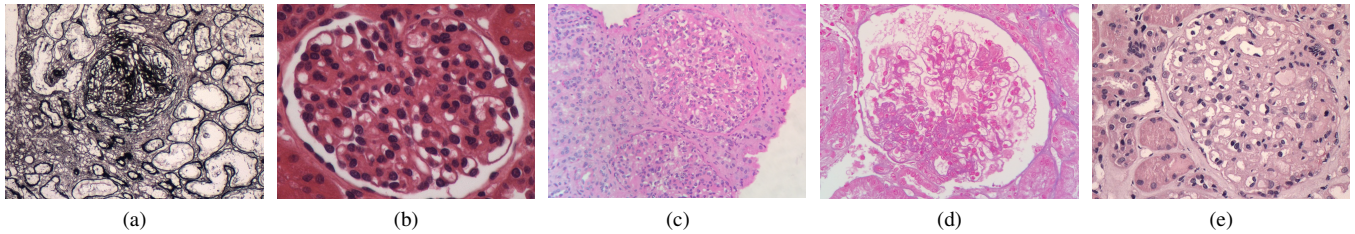


Fig. 1. Exemplos de glomerulos presentes na base. (a) Glomérulo com crescente corado em PAMS. (b) Glomérulo normal corado em H&E. (c) Glomérulo com hiper celularidade corado em PAS. (d) Glomérulo com esclerose corado em AZAN. (e) Glomérulo com membranosa corado em PS.

III. RESULTADOS E DISCUSSÃO

A. Classificação Crescente vs Normal

A Tabela I apresenta os resultados obtidos na classificação entre glomerulos normais e glomerulos com crescente, sem a aplicação de data augmentation no treinamento.

TABELA I
RESULTADOS DA CLASSIFICAÇÃO CRESCENTE VS NORMAL.

Rede	Acurácia	Precisão	Recall	F1-score
Xception	0.842 ± 0.025	0.814 ± 0.051	0.830 ± 0.065	0.819 ± 0.028
VGG16	0.909 ± 0.014	0.898 ± 0.026	0.892 ± 0.033	0.894 ± 0.018
Inception	0.813 ± 0.040	0.800 ± 0.090	0.790 ± 0.098	0.784 ± 0.031
ResNet50	0.958 ± 0.011	0.949 ± 0.017	0.955 ± 0.017	0.952 ± 0.012
MobileNet	0.841 ± 0.020	0.824 ± 0.029	0.805 ± 0.033	0.814 ± 0.024

A Tabela II apresenta os resultados obtidos na classificação entre glomerulos normais e glomerulos com crescente, com a aplicação de data augmentation no treinamento.

TABELA II
RESULTADOS DA CLASSIFICAÇÃO CRESCENTE VS NORMAL, COM DATA AUGMENTATION.

Rede	Acurácia	Precisão	Recall	F1-score
Xception	0.833 ± 0.028	0.818 ± 0.075	0.812 ± 0.116	0.804 ± 0.045
VGG16	0.915 ± 0.018	0.913 ± 0.022	0.889 ± 0.039	0.900 ± 0.023
InceptionV3	0.806 ± 0.046	0.785 ± 0.084	0.797 ± 0.181	0.769 ± 0.100
ResNet50	0.957 ± 0.010	0.948 ± 0.014	0.953 ± 0.016	0.951 ± 0.011
MobileNet	0.834 ± 0.030	0.819 ± 0.043	0.792 ± 0.058	0.804 ± 0.038

A rede ResNet50, considerando acurácia ($p < 0.001$), precisão ($p < 0.001$), recall ($p < 0.001$), e F1-score ($p < 0.001$), apresentou valores significativamente superiores aos de todas as outras redes. Mas a rede ResNet50 treinada com data augmentation não apresentou resultados significativamente superiores em relação à rede treinada sem data augmentation em nenhuma das métricas, acurácia ($p = 0.652$), precisão ($p = 0.699$), recall ($p = 0.562$), e f1-score ($p = 0.574$). Da mesma forma, as redes Xception, InceptionV3 e MobileNet também não obtiveram resultados significativamente superiores ao se utilizar dados ampliados via data augmentation.

A rede VGG16 com data augmentation também não apresentou resultados significativamente superiores em relação à rede sem data augmentation na acurácia ($p = 0.062$), no recall ($p = 0.676$), e no F1-score ($p = 0.167$). Contudo, a precisão ($p = 0.015$) do modelo com data augmentation se mostrou significativamente superior em relação à rede sem data augmentation.

De modo geral, pode ser afirmado que, no contexto dessa tarefa, o modelo com ResNet50 foi capaz de classificar os

glomerulos com crescente de forma mais efetiva que os demais, por conta de um melhor desempenho na comparação de todas as métricas avaliadas. No que se refere à aplicação de *data augmentation* no treinamento, na maioria dos casos, nenhum dos modelos obteve melhores resultados que o modelo inicial treinado sem o *data augmentation*, exceto o modelo com VGG16 em uma métrica específica, a precisão. Isso pode indicar que a forma de aplicação do *data augmentation* não foi suficiente para promover maior generalização aos modelos, o objetivo era fazer com que eles identificassem regiões da imagem relevantes à classificação.

B. Classificação Crescente vs Não-crescente

Ampliando o conjunto de dados com imagens de outras lesões, foi realizada a classificação entre glomerulos com crescente e sem crescente. Não foi aplicado *data augmentation* no treinamento, uma vez que não houveram resultados superiores na sua aplicação. Considerando o desbalanceamento do conjunto completo de imagens, foi feito uso de pesos por classe na entropia cruzada ponderada. Os resultados iniciais nesta condição foram acurácia de 0.891 ± 0.015 , precisão de 0.713 ± 0.066 , recall de 0.696 ± 0.072 , e f1-score de 0.699 ± 0.029 . Embora não seja possível uma comparação direta com os resultados da seção anterior, há uma indicação de que esta tarefa é mais difícil do que a anterior.

Uma forma de buscar melhorar os resultados seria focar nos exemplos mais difíceis durante o treinamento buscando posicionar melhor a fronteira de decisão. Isto foi realizado com a adição de um fator de modulação dinâmico adotando o ajuste dinâmico da entropia cruzada por Focal Loss [12]. Além disso, o Focal Loss também tem efeito no desbalanceamento. A Tabela III apresenta os resultados obtidos com uso de Focal Loss na classificação crescente vs não-crescente, sem data augmentation no treinamento.

TABELA III
RESULTADOS DA CLASSIFICAÇÃO CRESCENTE VS NÃO CRESCENTE COM FOCAL LOSS.

γ	α	Acurácia	Precisão	Recall	F1-score
2	0.500	0.898 ± 0.011	0.749 ± 0.041	0.668 ± 0.063	0.704 ± 0.038
2	0.750	0.899 ± 0.013	0.759 ± 0.050	0.659 ± 0.049	0.704 ± 0.037
5	0.750	0.898 ± 0.014	0.753 ± 0.045	0.658 ± 0.051	0.702 ± 0.043

Em relação à acurácia, a rede treinada com Focal Loss ($\gamma = 2$, $\alpha = 0.5$) se mostrou significativamente superior ao modelo que atribui pesos às classes ($p = 0.024$), o mesmo pode ser dito em relação à precisão ($p = 0.011$). Em relação as outras

métricas, recall ($p = 0.925$) e F1-score ($p = 0.368$), não pode ser dito que com Focal Loss ($\gamma = 2$, $\alpha = 0.5$) se mostrou significativamente superior ao modelo que atribui pesos.

Em relação à acurácia e precisão, as redes com outros parâmetros do Focal Loss também se mostraram significativamente superiores em relação ao modelo que atribui pesos às classes. No que diz respeito ao F1-score, não foi possível afirmar que os modelos com Focal Loss se apresentaram significativamente superiores. O modelo que atribui pesos se apresentou significativamente superior em relação ao modelo com Focal Loss ($\gamma = 2$, $\alpha = 0.75$), no que diz respeito à recall ($p = 0.022$), o mesmo aconteceu quando comparado ao modelo com Focal Loss ($\gamma = 5$, $\alpha = 0.75$), ($p = 0.010$).

Por fim, foi efetuado o teste de Wilcoxon para comparar os modelos que fizeram uso do Focal Loss, a fim de apontar, dentre as combinações de valores de γ e α utilizadas, qual proporcionou melhor desempenho na tarefa. Nenhum dos modelos se mostrou significativamente superior aos outros em nenhuma métrica avaliada.

Ao analisar os resultados desta tarefa, pode ser percebida uma queda na taxa de acurácia em relação à primeira tarefa de classificação, uma possível indicação de que as imagens das outras lesões, hiperplasmática, membranosa e esclerose, se assemelham mais às características dos crescentes do que as imagens dos glomérulos normais. Contudo, não pode ser atribuído somente a isso, o desbalanceamento entre as classes também é um fator a ser considerado. Na comparação com a ponderação de classe na entropia cruzada, o Focal Loss promoveu maior taxa de acerto na classificação, os testes indicaram resultados significativamente melhores de acurácia e precisão. No que diz respeito aos acertos da classe positiva, os crescentes, a abordagem com atribuição de pesos se mostrou significativamente superior aos modelos com Focal Loss com $\gamma = 2$ e $\alpha = 0.75$, e $\gamma = 5$ e $\alpha = 0.75$, a comparação entre os valores de recall justificam isso.

IV. CONSIDERAÇÕES FINAIS

Modelos e condições de aplicação de *deep learning* foram avaliados em duas tarefas de classificação de glomérulos com crescentes. Na classificação de imagens com glomérulos com crescente vs normais, a ResNet50 apresentou desempenho significativamente superior ao das demais nas medidas indicadas. A aplicação de algumas técnicas de *data augmentation* para incremento da base dados de treinamento, de maneira geral, não proporcionou melhoria no desempenho das redes nesta tarefa, apenas o modelo com VGG16 apresentou resultados melhores na precisão. Na classificação de glomérulos com crescente vs não crescente, modelos com ResNet50 apresentaram resultados superiores de acurácia e precisão no treinamento com Focal Loss, enquanto o recall e o f1-score tiveram valores superiores nos modelos com atribuição de pesos às classes.

Como possibilidades de trabalhos futuros, podem ser consideradas modificações na arquitetura dos modelos, a adição de camadas de regularização, bem como adicionar camadas entre a camada totalmente conectada e a camada de ativação, considerar outras maneiras de aplicação do *data augmentation*,

por exemplo, o uso de imagens produzidas sinteticamente por GANs (Generative Adversarial Networks). Uma pesquisa em andamento visa modificar a abordagem de treinamento, passando a efetuar-lo de maneira semi-supervisionada, como forma de promover uma melhor generalização aos modelos [13]. Também temos a intenção de inserir explicabilidade aos resultados da classificação, por meio da adoção de técnicas de interpretabilidade visual dos modelos [14].

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

- [1] P. Chagas, L. Souza, I. Araújo, N. Aldeman, A. Duarte, M. Angelo, W. L. dos Santos, and L. Oliveira, "Classification of glomerular hypercellularity using convolutional features and support vector machine," *Artificial Intelligence in Medicine*, vol. 103, p. 101808, Mar. 2020.
- [2] G. Bueno, M. M. Fernandez-Carobles, L. Gonzalez-Lopez, and O. Deniz, "Glomerulosclerosis identification in whole slide images using semantic segmentation," *Computer Methods and Programs in Biomedicine*, vol. 184, p. 105273, Feb. 2020.
- [3] A. B. Fogo and M. Kashgarian, *Diagnostic atlas of renal pathology*, 3rd ed. Philadelphia, PA: Elsevier, 2017.
- [4] L. Anguiano, R. Kain, and H.-J. Anders, "The glomerular crescent: triggers, evolution, resolution, and implications for therapy," *Current Opinion in Nephrology and Hypertension*, vol. 29, no. 3, p. 302, 2020.
- [5] G. O. Barros, B. Navarro, A. Duarte, and W. L. C. dos Santos, "PathoSpotter-k: A computational tool for the automatic identification of glomerular lesions in histological images of kidneys," *Scientific Reports*, vol. 7, no. 1, Apr. 2017.
- [6] J. N. Marsh, M. K. Matlock, S. Kudose, T.-C. Liu, T. S. Stappenbeck, J. P. Gaut, and S. J. Swamidass, "Deep learning global glomerulosclerosis in transplant kidney frozen sections," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2718–2728, Dec. 2018.
- [7] B. Ginley, B. Lutnick, K.-Y. Jen, A. B. Fogo, S. Jain, A. Rosenberg, V. Walavalkar, G. Wilding, J. E. Tomaszewski, R. Yacoub, G. M. Rossi, and P. Sarder, "Computational segmentation and classification of diabetic glomerulosclerosis," *Journal of the American Society of Nephrology*, vol. 30, no. 10, pp. 1953–1967, Sep. 2019.
- [8] E. Uchino, K. Suzuki, N. Sato, R. Kojima, Y. Tamada, S. Hiragi, H. Yokoi, N. Yugami, S. Minamiguchi, H. Haga, M. Yanagita, and Y. Okuno, "Classification of glomerular pathological findings using deep learning and nephrologist-AI collective intelligence approach," *International Journal of Medical Informatics*, vol. 141, p. 104231, Sep. 2020.
- [9] P. Chagas, L. Souza, I. Pontes, R. Calumby, M. Angelo, A. Duarte, W. dos Santos, and L. Oliveira, "Deep-learning-based membranous nephropathy classification and monte-carlo dropout uncertainty estimation," in *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*. Porto Alegre, RS, Brasil: SBC, 2021, pp. 257–268.
- [10] J. M. C. Rehem, W. L. C. dos Santos, A. A. Duarte, L. R. de Oliveira, and M. F. Angelo, "Automatic glomerulus detection in renal histological images," in *Medical Imaging 2021: Digital Pathology*, J. E. Tomaszewski and A. D. Ward, Eds. SPIE, Feb. 2021.
- [11] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, pp. 1–54, 2019.
- [12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [13] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 3, no. 1, pp. 1–130, Jan. 2009.
- [14] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020.