

Classificação de Imagens Histológicas de Glomérulos Renais com Podocitopatia

George Oliveira Barros (190052911)
george.barros@ifgoiano.edu.br
Prof. Dr. Teófilo de Campos

Departamento de Ciência da
Computação
Universidade de Brasília
Campus Darcy Ribeiro, Asa Norte
Brasília-DF, CEP 70910-900, Brazil,

Abstract

Este artigo descreve a implemetação e comparação de alguns modelos de redes neurais convolucionais para classificação de imagens histológicas renais de glomérulos como imagens com lesão podocitária ou imagens sem lesão. Este trabalho é o projeto final da disciplina de visão computacional, ministrada pelo professor Teófilo de Campos no CIC/UnB. Os melhores resultados obtidos chegaram a 93% de acurácia. Muitas abordagens e melhorias podem ser realizadas, contudo o resultado foi considerado satisfatório, tendo em vista o objetivo dos experimentos, que foi o de comparar arquiteturas residuais e densas com diferentes camadas de profundidade em um ambiente bem restrito de treino, utilizando transfer learning com pesos do dataset Imagenet.

1 Introdução

Aprendizado profundo (deep learning) é uma abordagem de aprendizado de máquina que tem crescido muito desde a última década. As técnicas empregadas pelos algoritmos, basicamente, se baseiam em redes neurais convolucionais organizadas com múltiplas camadas ocultas [1]. Em termos gerais, essas redes também podem ser vistas como encadeamentos de funções lineares seguidas de uma ou mais funções de ativação (não lineares).

Entre as áreas mais beneficiadas pelo desenvolvimento das redes neurais profundas, encontra-se a visão computacional, que tem como objetivo tornar possível a tomada de decisão e extração de informações a partir de imagens e vídeos [2]. Os domínios de aplicação da visão computacional são muitos, no entanto, a análise de imagens médicas talvez tenha sido uma das aplicações que ganhou mais notoriedade em relação à utilização dos algoritmos de aprendizado profundo, seja para classificação, detecção, segmentação ou até mesmo criação de imagens sintéticas [3], [4] e [5].

As aplicações da visão computacional têm impactado a prática médica e influenciando principalmente a qualidade dos diagnósticos. Entre os distintos temas de pesquisa originados a partir desta parceria, destaca-se a histopatologia digital, considerada como uma das maiores evoluções da medicina moderna [6]. A histopatologia digital pode ser compreendida a partir dos conceitos de duas outras áreas, a histologia e a histopatologia. A histologia é o estudo da anatomia microscópica de tecidos de organismos. A histopatologia, por sua vez, é a análise microscópica de seções histológicas com o objetivo da diferenciação entre tecidos

biológicos saudáveis e doentes [4] [5]. Logo, a histopatologia digital é a análise de seções de tecido biológico através de sistemas computacionais.

Existe uma série de trabalhos com imagens histológicas, sobretudo com foco no estudo de câncer. Contudo, ainda há carência de pesquisas relacionadas a algumas doenças específicas, como é o caso das podocitopatias. As podocitopatias são uma família de doenças que afetam os podócitos de um glomérulo, que por sua vez é uma estrutura biológica contida no rim.

Tendo em vista a importância do estudo das podocitopatias e a inexistência de trabalhos sobre essas doenças, o objetivo deste trabalho foi comparar algumas arquiteturas de rede tradicionais para classificar imagens de glomérulos renais como: “com” ou “sem” podocitopatia.

2 Revisão de literatura

A análise de imagens histológica consiste no padrão ouro para o diagnósticos de patologias. Adicionalmente, a taxa de sobrevivência renal tende a ser melhor com avaliação histológica por biópsia renal do que sem biópsia renal. Deste modo a utilização de técnicas de apoio ao diagnóstico médico por imagem são muito importantes. A seguir alguns trabalhos que aplicam redes neurais profundas em imagens histológicas renais de glomérulos ou de lâminas de tecido em geral.

[6] utilizam CNNs para extrair características de imagens histológicas de glomérulos de camundongos e classificar em relação às diferenças entre animais de genótipo diferentes. Os resultados foram uma distribuição visual clara das classes de amostras do dataset.

Em [7] o objetivo foi prever a perda de função renal de pacientes com doença renal crônica, tarefa realizada a partir da taxa de filtragem de sangue pelo glomérulo. Os resultados obtidos são de uma taxa de erro de 17,11%.

O trabalho de [8] tratou-se de uma proposta de modelo baseado em técnicas em CNNs para diferenciar tecidos renais com glomeruloesclerose e tecidos normais. Os autores utilizam a arquitetura de rede VGG e um dataset de 48 lâminas de tecido. Os resultados obtidos foram de 81% de acurácia na identificação de glomérulos não esclerosados e 60% em relação a glomérulos esclerosados.

[9] e [10], por sua vez, realizaram a segmentação de glomérulos em imagens de lâmina renal. [9] utiliza as redes Faster R-CNN e Inception ResNet e atinge um resultado de mais de 90% de acurácia, com um dataset de 33 mil imagens histológicas. [10] aplica técnicas clássicas de processamento de imagens antes de realizar a segmentação com a CNN em um dataset de 40 lâminas de tecido, alcançando uma acurácia de 96% para segmentar os glomérulos.

Entre os trabalhos mais recentes encontrados, até o presente momento, [11] realiza três tarefas por meio da arquitetura de rede neural convolucional U-net. Os autores detectam o glomérulo renal, classificam quanto ao tipo de lesão e ainda identificam e quantificam as células internas do glomérulo com uma precisão e 93,1% e recall de 94,9%. [12] apresentam um modelo baseado em DNN para classificação de esclerose global, resultando em um desempenho de 0,986 de área sob a curva roc.

Até o momento não foi encontrado trabalhos com foco específico na classificação de podocitopatias. Isso se dá pelo fato da doença ter uma incidência bem menor do que as citadas anteriormente e provavelmente haver poucos patologistas para rotulação das imagens com podocitopatias.

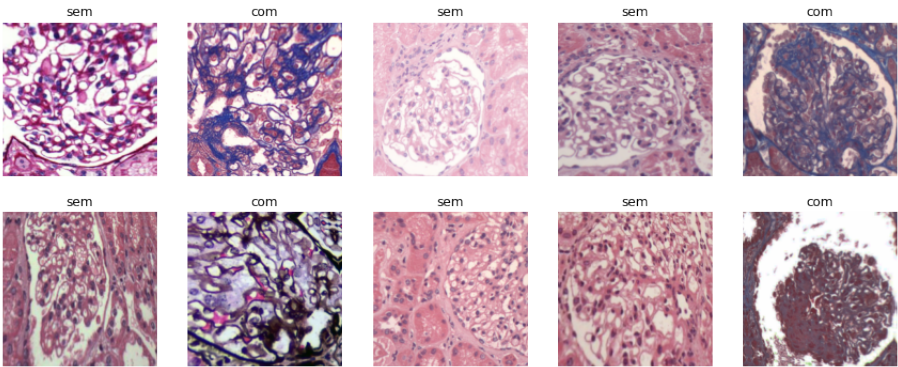


Figure 1: Exemplos de imagens que compõem o dataset.

3 Metodologia

A metodologia utilizada para o treinamento das redes e validação dos resultados se dividiu nas seguintes etapas: Preparação do dataset, Escolha das arquiteturas de rede, Treinamento e ajuste, Avaliação. Foi utilizado transfer learning. Todas as redes foram iniciadas a partir dos pesos das arquiteturas de rede pré-treinadas com o dataset Imagenet.

3.1 Preparação do dataset

O dataset utilizado consiste de um conjunto de 376 imagens de glomérulos renais rotulados como: "sem lesão podocitária" ou "com lesão podocitária". Do conjunto total, 176 são imagens com lesão e 200 sem lesão. As imagens foram anotadas por dois patologistas da Fundação Osvaldo Cruz-BA. As imagens estão coradas em três diferentes padrões de corantes: HE, PAS e Tricômio. A Figura 1 apresenta alguns exemplos de imagens que compõem o dataset. Não há um padrão de resolução e dimensão das imagens, pois foram adquiridas de dispositivos distintos. Logo, foi necessário normalizá-las antes do input das redes. Como foi utilizado os pesos da Imagenet, isso foi feito automaticamente pela biblioteca Fastai [16]. Adicionalmente, foram aplicadas uma série de técnicas de data augmentation, como: flip vertical e horizontal, rotação de 90 e 180 graus, aumento de brilho e zoom aleatório de no mínimo 10% e no máximo 30%. A Figura 2 mostra os resultado do data augmentation em uma imagem específica. Essas técnicas foram as selecionadas pelo fato de, além de ampliarem o conjunto de amostras, também não descaracterizarem as imagens. O aumento de dados foi necessário para evitar problemas como, underfitting, por exemplo.

3.2 Arquiteturas de rede utilizadas

As arquiteturas de rede utilizadas nas comparações foram versões de redes residuais e redes densas: Resnet34, Resnet50, Renet101, Densenet121 e Densenet201. O objetivo foi avaliar tipos diferentes de redes profundas em diferentes níveis de profundidade.

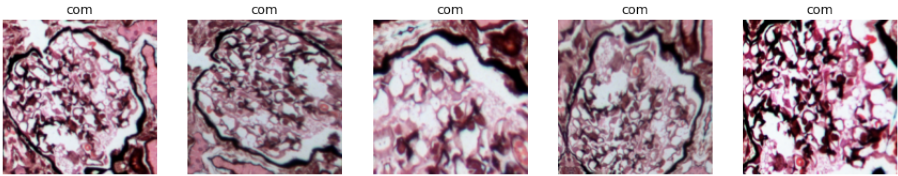


Figure 2: Resultado do data augmentation.

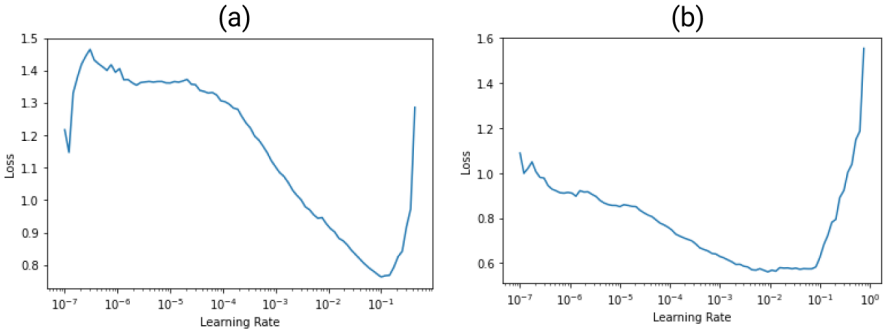


Figure 3: (a) Gráfico para encontrar o learning rate na camada recém adicionada à rede. (b) Encontrar o learning rate para treinar a rede completa.

3.3 Treinamento e ajuste das redes

Para todas as redes utilizadas foi realizado como descrito a seguir:

O treinamento foi realizado com uma divisão simples de 75% dos dados para treino e 25% para teste (de maneira automática). O mesmo conjunto de treino e teste foi utilizado para todas as redes. O método de ajuste dos parâmetros seguiu a aula 05 do fastbook (literatura referência da disciplina). Como foi realizado transfer learning, inicialmente foi necessário retirar as últimas camadas das redes e retrainá-las para o novo dataset, apenas um rápido ajuste. Por padrão a biblioteca fast ai realiza uma única época de treino para o ajuste dos pesos. Porém, neste trabalho foram executadas 5 épocas. Antes do ajuste, contudo, fez-se necessário encontrar a taxa de aprendizado mais adequada, para conseguir um treinamento otimizado.

Após o ajuste da última camada (recém adicionada à rede), realizou-se o "descongelamento" da rede, que se trata de habilitar as demais camadas da rede. Após isso, novamente procurou-se a taxa de aprendizado mais adequada (agora para rede inteira). A Figura 3 mostra o gráfico de taxa de aprendizado da Resnet50, inicialmente para a última camda de rede e também para o treinamento otimizado da rede inteira. Neste método, proposto por [4], escolhe-se uma taxa de aprendizado bem pequena e aplica-a a um minilote, em seguida calcula-se a perda e dobra-se a taxa de aprendizado para outro minilote e repete-se o processo. Isto ocorre até a perda aumentar (sinal que o tamanho da taxa de aprendizado foi muito grande). Por fim, escolhe-se a última taxa onde o erro ainda foi reduzido.

Em seguida, a próxima etapa foi a de treinar a rede inteira. Os demais hiperprâmetros

como tamanho de minilote e função de perda foram mantidos automáticos para todas as redes. No caso da função de perda a utilizada em todas as redes foi a entropia cruzada binária, bastante adequada, por ser um problema clássico de classificação binária. A função de ativação, no final da rede, foi a softmax, que normaliza os dados entre os valores 0 e 1.

Cada rede foi treinada com um número específico de épocas. Aós vários testes e análise dos gráficos de perda no treino e teste no decorrer das ápodas pôde-se observar a partir de quando as redes entravam em overfitting. A Figura 4, trata-se dos gráficos de treino da RESNET101 e exemplifica isso. A partir do minilote 32 a perda do conjunto de validação (teste) começa aumentar, enquanto a perda do conjunto de treino continua caindo.

3.4 Avaliação

As métricas utilizadas para avaliar os modelos foram a perda, ainda no treinamento e após treinamento, a acurácia, precisão, recall e f1-score, ambas métricas obtidas a partir da matriz de confusão dos modelos de classificação, obtida por sua vez, através da avaliação do conjunto de teste. A equações de cada métrica encontram-se em [15].

4 Resultados

A Tabela 1 apresenta os resultados de F1-score, acurácia, precisão e recall de cada uma dos modelos resultantes do transfer learning. Adicionalmente, a Tabela 1 apresenta o número de épocas que o modelo foi treinado. Como pode ser observado em negrito, com base no conjunto de treino/teste utilizado no treinamento dos modelos, a Resnet101 foi a que apresentou os melhores resultados. O f1-score merece um destaque especial. Percebe-se que foi o modelo com o melhor equilíbrio entre acertos entre as imagens com lesão e imagens sem lesão.

Modelo	Acc.(%)	Prec.(%)	Recc.(%)	F1-score(%)	Pesos iniciais	Épocas
Resnet34	0,86	0,79	0,94	0,86	Imagenet	15
Resnet50	0,89	0,84	0,89	0,86	Imagenet	13
Resnet101	0,93	0,89	0,94	0,92	Imagenet	7
Densenet121	0,87	0,82	0,86	0,84	Imagenet	14
Densenet201	0,92	0,91	0,89	0,90	Imagenet	14

Table 1: Comparação dos resutados obtidos com todas os modelos

5 Conclusões e Discussão

Dado o número pequeno de imagens e o fato dos testes terem sido realizados com apenas um conjunto de treino e teste, em ambitos gerais, não é possível afirmar de fato, qual é o melhor modelo. No entanto, os resultados indicam, com base apenas nos dados utilizados, que o modelo Resnet101 foi o que forneceu os melhores resultados com um menor número de épocas de treino.

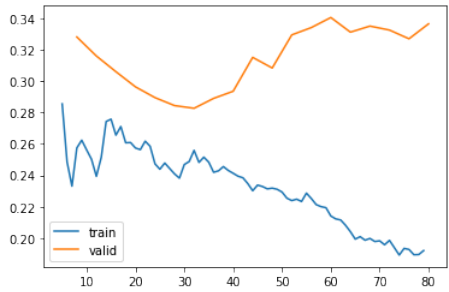


Figure 4: (a) Gráfico para encontrar o learning rate na camada recém adicionada à rede. (b) Encontrar o learning rate para treinar a rede completa.

6 Trabalhos Futuros

Há muitas abordagens de melhoria que podem ser implementadas neste trabalho. Inicialmente, a abordagem de validação dos modelos pode ser melhorada com a adoção do método de validação cruzada k-fold. O dataset também deve ser melhorado. Com apenas 376 imagens os resultados se tornam bastante suscetíveis a resultados enganosos. Além da adição de mais amostras, os dados devem ser validados por um número maior de patologistas. Tendo em vista a complexidade das lesões podocitárias, pode haver discordância de diagnóstico entre patologistas.

Em relação a mais experimentos, um bom comparativo pode ser a realização do treinamento das redes from scratch (do zero), sem transfer learning. Outra fragilidade ainda, foi a não realização do ajuste dos hiperparâmetros da rede. Apesar da ferramenta utilizada realizar escolhas automáticas dos hiperparâmetros, só a realização de testes variando esses valores para garantir as melhores escolhas possíveis.

Por fim, como trabalho futuro, pode ser realizado a classificação dos tipos de lesões podocitárias que são: hiperplasia, degeneração e hipertrofia, tornando o problema em uma classificação multiclasse. Ou ainda, após anotação da localização das lesões, realizar a segmentação e classificação das lesões podocitárias.

References

- [1] Goodfellow et al., "Deep Learning," MIT Press, 2016, Disponível em: <http://www.deeplearningbook.org>, acessado 01 de setembro de 2019.
- [2] Pedrini, H., Schwartz, W. R. (2008). Análise de imagens digitais: princípios, algoritmos e aplicações. CENGAGE - UM LIVRO. Retrieved from <https://books.google.com.br/books?id=13KAPgAACAAJ>
- [3] Irshad, H. et al., Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential. Ieee Reviews In Biomedical Engineering, Vol. 7, 2014.
- [4] Belsare, A. D.; Mushrif, M. M. Histopathological image analysis using image pro-

- cessing techniques: an overview. Signal & image processing : an international journal (SIPIJ) vol.3. 2012.
- [5] Lei He, L.; et al. Histology image analysis for carcinoma detection and grading. Computer methods and programs in biomedicine. v 107 p. 538–556. 2012.
- [6] Leslie N. Smith. Cyclical Learning Rates for Training Neural Networks. arXiv:1506.01186v6 [cs.CV] 4 Apr 2017.
- [7] Ledbetter, D., Ho, L., & Lemley, K. (2017). Prediction of Kidney Function from Biopsy Images Using Convolutional Neural Networks.
- [8] Ginley2017: Ginley, B. G., Emmons, T., Lutnick, B., Urban, C. F., Segal, B. H., & Sarder, P. (2017). Computational detection and quantification of human and mouse neutrophil extracellular traps in flow cytometry and confocal microscopy. Scientific Reports, 7(1), 1–11. <https://doi.org/10.1038/s41598-017-18099-y>
- [9] Marsh, J. N., Stappenbeck, T. S., Liu, T.-C., Swamidass, S. J., Gaut, J. P., Kudose, S., & Matlock, M. K. (2018). Deep Learning Global Glomerulosclerosis in Transplant Kidney Frozen.
- [10] Lutnick, B., Ginley, B., Govind, D., McGarry, S. D., Laviolette, P. S., Yacoub, R., ... Sarder, P. (2019). Iterative annotation to ease neural network training: Specialized machine learning in medical image analysis. Nature Machine Intelligence, 1(December), 112–119. <https://doi.org/arXiv:1812.07509v1>
- [11] Sheehan S, Mawe S, Cianciolo RE, Korstanje R, Mahoney JM. Detection and Classification of Novel Renal Histologic Phenotypes Using Deep Neural Networks. Am J Pathol. 2019 Sep;189(9):1786-1796. doi: 10.1016/j.ajpath.2019.05.019. Epub 2019 Jun 18. PMID: 31220455; PMCID: PMC6723211.
- [12] Fukayama, M., Uozaki, H., Shimamoto, K., Ohe, K., Kawazoe, Y., Yamaguchi, R., & Shintani-Domoto, Y. (2018). Faster R-CNN-Based Glomerular Detection in Multistained Human Whole Slide Images. Journal of Imaging, 4(7), 91. <https://doi.org/10.3390/jimaging4070091>
- [13] Eiichiro Uchino, Kanata Suzuki, Noriaki Sato, Ryosuke Kojima, Yoshinori Tamada, Shusuke Hiragi, Hideki Yokoi, Nobuhiro Yugami, Sachiko Minamiguchi, Hironori Haga, Motoko Yanagita, Yasushi Okuno. Classification of glomerular pathological findings using deep learning and nephrologist–AI collective intelligence approach. International Journal of Medical Informatics doi: 10.1016/j.ijmedinf.2020.104231, 2020.
- [14] Zeng, C., Nan, Y., Xu, F., Lei, Q., Li, F., Chen, T., ... Liu, Z. (2020). Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning. <https://doi.org/10.1002/path.5491>
- [15] Russel, Stuart, J. (Stuart Jonathan), 1962 - Inteligência artificial/Stuart Russell - tradução Regina Célia Smille. - Rio de Janeiro: Elsevier, 2013.
- [16] Jeremy Howard. Image Classification. Fastbook. Disponível em: https://github.com/fastai/fastbook/blob/master/05_pets_reeds.ipynb. Acessado em 07 de dezembro de 2020.