

# Geoestatísticas e *Deep Features* para Diferenciar Glomeruloesclerose Segmentar e Focal de Doença de Lesões Mínimas em Imagens de Biópsia Renal

Justino Duarte Santos<sup>1</sup>, Rodrigo M. S. Veras<sup>1</sup>, Romuere R. V. Silva<sup>2</sup>

<sup>1</sup> Universidade Federal do Piauí – UFPI  
Departamento de Computação Teresina-PI, Brasil

<sup>2</sup> Universidade Federal do Piauí  
Curso de Sistemas de Informação Picos-PI, Brasil

justinoduarte@gmail.com, {rveras, romuere}@ufpi.edu.br

**Abstract.** *Chronic kidney diseases arise from acute or intermittent pathologies that have not been adequately treated, such as minimal change disease (MCD) and focal segmental glomerulosclerosis (FSGS). The accurate identification of these two diseases is of paramount importance, because their treatments and prognoses are different. Thus, we propose a method that is capable of differentiating MCD from FSGS based on images from pathological examinations. In the proposed method, we use four pre-trained convolutional neural networks and geostatistical functions to extract image features. We selected 94 features based on mutual information criteria, and in the classification step, we used a random forest classifier. The proposed method obtained an accuracy of 94.3% and Kappa index of 87.9%, confirming that our method is very promising.*

**Resumo.** *As doenças renais crônicas surgem de patologias agudas ou intermitentes quando não tratadas adequadamente, como a doença de lesões mínimas (DLM) e glomeruloesclerose segmentar e focal (GESF). A identificação precisa dessas duas doenças é de suma importância, pois seus tratamentos e prognósticos são diferentes. Dessa forma, propomos um método capaz de diferenciar a DLM de GESF a partir de imagens de exames patológicos. No método proposto, usamos quatro redes neurais convolucionais pré-treinadas e funções geoestatísticas para extrair características de imagem. Seleccionamos 94 características com base em critérios de informação mútua e, na etapa de classificação, usamos um classificador random forest. O método proposto obteve acurácia de 94,3% e índice Kappa de 87,9%, confirmando que nosso método é bastante promissor.*

## 1. Introdução

Diversas anormalidades histológicas podem levar a uma síndrome nefrótica (SN), mas se destacam como sua causa a Doença de Lesões Mínimas e a Glomeruloesclerose Segmentar e Focal. Nas crianças, a DLM é a causa da SN em 90% dos pacientes, enquanto nos adultos, doenças como GESF e DLM são a causa da SN em 70% dos casos. Na população adulta, a GESF é a principal causa de SN em vários países. [Moura et al. 2015].

É essencial diferenciar DLM de GESF, pois os tratamentos e a resposta à eles são distintos, bem como os prognósticos destas duas doenças [Moura et al. 2015]. Vista essa importância, este trabalho propõe um método computacional de classificação capaz de distingui-las a partir de imagens de biópsia.

## 2. Trabalhos Relacionados

Dentre os principais trabalhos focados em tarefas de classificação em imagens de rim, podemos encontrar o trabalho de [Barros et al. 2017], no qual foi proposto um sistema computacional para detectar Lesão Glomerular Proliferativa (PGL). Utilizaram o *k-nearest neighbor* (KNN)[Aha et al. 1991] como classificador. Já Chagas et al. [Chagas et al. 2019], além da detecção de PGL, também faz classificação de subtipos específicos de PGL: endocapilar, mesangial e ambas. Eles construíram um arquitetura baseada em CNN para extrair características das imagens, classificaram usando uma SVM [Cortes and Vapnik 1995].

Os autores [Ginley et al. 2018] propuseram uma abordagem para definir a progressão estrutural de nefropatia diabética em glomérulos. O método extraiu características baseadas na análise de textura, morfologia e distâncias.

Sheehan e Korstanje [Sheehan and Korstanje 2018] desenvolveram um método semi-automático para identificar o glomérulo e quantificar informações fenotípicas. O *Random Forest* (RF) [Ho 1995] foi empregado como classificador.

Marsh et al. [Marsh et al. 2018] trabalharam com um modelo de aprendizado profundo que classifica glomérulos com ou sem esclerose. A abordagem proposta ajustou a CNN VGG-16 [Simonyan and Zisserman 2014].

Nos trabalhos acima encontramos o uso de várias técnicas aplicadas em imagens de rim. Aplicamos um conjunto delas para para diferenciar GESF e DLM. Não encontramos bases de dados nem outros trabalhos que proponham métodos para isso. Portanto estas são as principais contribuições deste.

## 3. Materiais e métodos

Nesta seção, apresentamos um método para diferenciar imagens de biópsia renal com GESF ou DLM. Realizamos experimentos usando características de textura como as de Haralick [Haralick et al. 1973] e geoestatísticas, e *deep features* de CNNs pré-treinadas.

### 3.1. Método proposto

A Figura 1 dá uma visão do método proposto. Ele contém quatro etapas: pré-processamento, extração de características, seleção de características e classificação. Além das técnicas apresentadas em cada etapa na imagem, outras também foram avaliadas, essas etapas serão detalhadas a seguir.

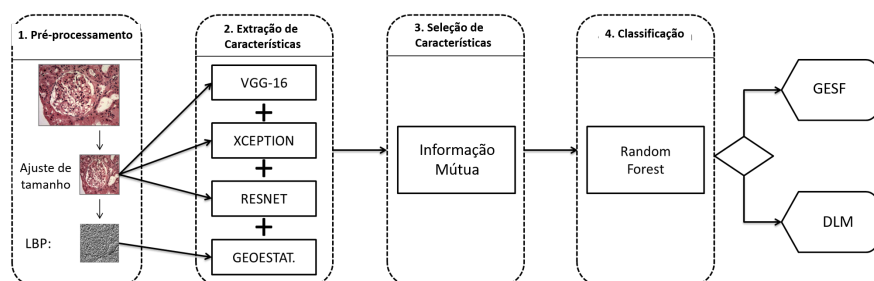


Figura 1. Fluxograma com etapas do método proposto.

#### 3.1.1. pré-processamento

Na etapa de pré-processamento, as imagens foram redimensionadas para as dimensões de entrada padrões das CNNs utilizadas. Para extrair informações de textura com as funções ge-

**Tabela 1. Sumário de características avaliadas neste trabalho**

Tipo	Método	Características	Numero de atributos
	Geoestat.	Semivariograma, semimadograma, covariograma e correlograma	528
Textura	Haralick	Contraste, dissimilaridade, homogeneidade, ASM e correlação	660
	VGG-16		4.096
Deep features	VGG-19	Extração da saída da penúltima camada da arquitetura.	4.096
	Xception		2.048
	ResNet50		2.048
<b>TOTAL</b>			<b>13.476</b>

oestatísticas, as imagens também foram redimensionadas ( $299 \times 299$ ) e então representadas na forma de LBPs [Ojala et al. 1996] para cada canal de cor. Para extrair as características de Haralick, as imagens foram usadas em seus tamanhos originais e suavizadas por meio da aplicação de um filtro mediano com uma janela de  $9 \times 9$  para reduzir os efeitos causados por eventuais ruídos.

### 3.1.2. Extração de Características

Neste trabalho, usamos quatro funções geoestatísticas [Isaaks et al. 1989]: semivariograma, semimadograma, covariograma e correlograma. Essas funções sintetizam o nível de associação entre os valores coletados em função da distância e da direção [Silva et al. 2004].

Computamos as quatro funções supracitadas, para cada canal de cor (RGB) usando 4 direções ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ ) e 11 distâncias em pixels, sendo 6 fixas (1, 2, 5, 10, 15, 20) e 5 proporcionais às dimensões da imagem (1.25%, 2%, 5%, 10%, e 20%).

Ainda de textura, avaliamos também as seguintes características de Haralick [Haralick et al. 1973]: contraste, dissimilaridade, homogeneidade, segundo momento angular (ASM) e correlação. Elas são calculadas com base na *gray-level co-occurrence matrix* (GLCM). A textura também é tomada em função de direção e distância. Utilizamos as mesmas definidas para geoestatísticas.

Nós aplicamos TL na extração de características extraído o vetor de saída da penúltima camada de quatro CNNs pré-treinadas na base da ImageNet [Russakovsky et al. 2015]: VGG-16, VGG-19 [Simonyan and Zisserman 2014], Xception [Chollet 2017] e ResNet50 [He et al. 2015].

A tabela 1 mostra o sumário dos vetores de características extraídos. Além de avaliá-los individualmente também geramos descritores combinados a partir deles. Assim ao todo tivemos 63 vetores.

Nas seções seguintes poderemos usar as seguintes abreviações: v16=VGG-16, v19=VGG-19, xce=Xception, rsnet=ResNet50, hrlk:Haralick and geo=geoestatísticas.

### 3.1.3. Seleção de características

Realizamos o processo de seleção de características para cada um dos 63 vetores de obtidos na etapa anterior, classificando os atributos no vetor em ordem decrescente de relevância. Avaliamos a estatística F de análise de variância (ANOVA-F) e informação mútua (IM) [Cover and Thomas 1991] para calcular a relevância de cada atributo.

Procedemos uma abordagem incremental de seleção para definir a dimensão do vetor

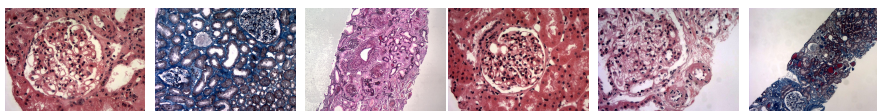


Figura 2. Exemplos de imagens da base. as três imagens da esquerda são GESF; as três imagens da direita são DLM.

a ser levado para classificação. Tomamos as características ordenadas (por ANOVA-F e IM), começamos com a mais relevante e levamos para classificação, em seguida adicionamos uma a uma as demais, e novamente as suprimos aos classificadores.

### 3.1.4. Classificação

Buscamos na abordagem incremental, na etapa de classificação, a quantidade de características do vetor ordenado que proporcionasse o melhor resultado, com o menor número de atributos.

No método proposto nós usamos o *Random Forest*, mas avaliamos também outros dois classificadores SVM e KNN.

Após a análise dos resultados das classificações, selecionamos os dez melhores modelos, e para eles executamos uma busca extensiva de hiper-parâmetros de configuração do classificador que aprimorassem os resultados.

### 3.2. Métricas de Avaliação

Utilizamos a técnica de validação cruzada *k-fold*, que distribui aleatoriamente as instâncias da base em *k* subconjuntos (*folds*) mutuamente exclusivos. A classificação acontece *k* vezes, em cada uma um *fold* é usado para testar e os demais para treinar. Então se calcula uma média de cada métrica considerada.

As métricas calculadas neste trabalho foram acurácia, kappa [Cohen 1960], precisão e *recall*. Nós usamos o Kappa como a principal métrica de avaliação. Ela nos dá um valor que representa o grau de concordância entre as classificações nominais realizadas por dois avaliadores - no caso presente, as previstas pelo classificador e as anotadas pelo patologista.

### 3.3. Base de Imagens

A base de imagens utilizada neste trabalho é composta por 87 imagens, sendo 33 da classe GESF e 54 de DLM. A Figura 2 mostra exemplos dessas imagens. Em muitos casos, observa-se heterogeneidade visual entre imagens pertencentes a uma mesma classe, bem como semelhança entre imagens em classes distintas, o que torna mais desafiadora a classificação.

## 4. Resultados e Discussão

Avaliamos 378 cenários ( $63$  vetores de características  $\times$   $2$  seletores de atributos  $\times$   $3$  classificadores). E ainda executamos o método incremental de seleção de atributos.

As Tabela 2 mostra os 5 melhores resultados de classificação do classificador RF. Todos eles foram obtidos usando IM para selecionar as características (a coluna #F indica o número de atributos selecionados). Com o KNN e SVM os resultados foram inferiores e não serão mostrados.

Tabela 2. Cinco melhores kappa alcançados usando RF (resultados obtidos com o seletor IM).

Vetor	#F	acurácia	kappa	precisao	recall
v16+v19+xce	62	<b>93.3±6.5</b>	<b>85.5±14.1</b>	91.9±7.2	<b>98.2±3.6</b>
v16+v19+ hrlk+geo	216	<b>93.3±6.5</b>	<b>85.5±14.1</b>	91.9±7.2	<b>98.2±3.6</b>
v16+v19+geo	434	<b>93.3±6.5</b>	<b>85.5±14.1</b>	91.9±7.2	<b>98.2±3.6</b>
v16+xce+geo	61	93.2±4.1	85.4±9.0	93.3±5.9	96.4±4.5
v16+xce+ rsnet+geo	94	93.2±4.1	85.2±8.9	91.6±4.9	<b>98.2±3.6</b>

- Os melhores resultados estão em negrito.

Tabela 3. Comparação dos resultados da classificação antes e depois do ajuste do hiperparâmetros.

Vector	Métrica	Antes	Depois
v16+xce+geo	acurácia	93.2±4.1%	<b>94.2±3.5%</b>
	kappa	85.4±9.0%	<b>87.4±7.8%</b>
	precisão	93.3±5.9%	<b>93.4±5.9%</b>
	recall	96.4±4.5%	<b>98.2±3.6%</b>
v16+xce+rsnet+geo	acurácia	93.2±4.1%	<b>94.3±3.5%</b>
	kappa	85.2±8.9%	<b>87.9±7.4%</b>
	precisão	91.6±4.9%	<b>94.7±4.3%</b>
	recall	<b>98.2±3.6%</b>	96.4±4.5%

- Os melhores resultados estão em negrito.

Para melhorar os resultados, avaliamos 211.200 configurações de hiperparâmetros para o RF (variando o número de árvores, profundidade etc.) em busca da melhor configuração. Não houve mudanças para os três melhores resultados da tabela 2); mas a Tabela 3 mostra o comparativo para os dois vetores subsequentes. Percebe-se que quase todas as métricas melhoraram.

Finalmente, após o ajuste de hiperparâmetros, definimos o melhor vetor de características: v16+xce+rsnet+geo. Dele são selecionadas 94 características baseadas na IM e classificadas com RF. Analisamos a origem de cada uma das características selecionadas. Nove delas são de geoestatísticas. É notável que, tendo apenas 528 atributos, as geoestatísticas tiveram uma taxa significativa de seleção quando comparada com os grandes vetores das CNNs.

## 5. Conclusão e Trabalhos Futuros

A caracterização da imagem por meio de um único tipo de descritor demonstra pouco poder de representação. Por outro lado, a criação de um descritor de natureza mista, como *deep features* e textura, são capazes de extrair atributos da imagem de maneira mais rica.

Como trabalhos futuros pretendemos avaliar métodos que utilizam TL, mas no formato de *fine tuning* em redes neurais convolucionais. Essa técnica faz um ajuste fino nos pesos de camadas específicas CNN.

## Referências

- Aha, D. W., Kibler, D., and Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1):37–66.
- Barros, G. O., Navarro, B., Duarte, A., and Dos-Santos, W. L. (2017). Pathospotter-k: A computational tool for the automatic identification of glomerular lesions in histological images of kidneys. *Scientific reports*, 7:46769.
- Chagas, P., Souza, L., Araújo, I., Aldeman, N., Duarte, A., Angelo, M., dos Santos, W. L., and Oliveira, L. (2019). Classification of glomerular hypercellularity using convolutional features and support vector machine. *arXiv preprint arXiv:1907.00028*.

- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *arXiv preprint*, pages 1610–02357.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.
- Cover, T. M. and Thomas, J. A. (1991). Entropy, relative entropy and mutual information. *Elements of information theory*, 2:1–55.
- Ginley, B. G., Tomaszewski, J. E., Jen, K.-Y., Fogo, A., Jain, S., and Sarder, P. (2018). Computational analysis of the structural progression of human glomeruli in diabetic nephropathy. In *Proceedings of SPIE Medical Imaging*, volume 10581, pages 105810A–1–105810A–6.
- Haralick, R. M., Shanmugam, K., et al. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 3(6):610–621.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282.
- Isaaks, E., Srivastava, R., and (Firm), K. (1989). *An introduction to Applied Geostatistics*. Oxford University Press.
- Marsh, J. N., Matlock, M. K., Kudose, S., Liu, T., Stappenbeck, T. S., Gaut, J. P., and Swamidass, S. J. (2018). Deep learning global glomerulosclerosis in transplant kidney frozen sections. *IEEE Transactions on Medical Imaging*, 37(12):2718–2728.
- Moura, L. R., Franco, M. F., and Kirsztajn, G. M. (2015). Minimal change disease and focal segmental glomerulosclerosis in adults: response to steroids and risk of renal failure. *Brazilian Journal of Nephrology*, 37(4):475–480.
- Ojala, T., Pietikäinen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.
- Sheehan, S. M. and Korstanje, R. (2018). Automatic glomerular identification and quantification of histological phenotypes using image analysis and machine learning. *American Journal of Physiology-Renal Physiology*, 315(6):F1644–F1651.
- Silva, A. C., Carvalho, P. C. P., and Gattass, M. (2004). Analysis of spatial variability using geostatistical functions for diagnosis of lung nodule in computerized tomography images. *Pattern Analysis and Applications*, 7(3):227–234.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.