

Um Pipeline de Visão Computacional para Análise de Achados Hiperecogênicos em Ultrassonografia Renal

Jefferson S. dos Anjos¹, Amaro A. de Lima¹, Gabriel M. Araújo¹,
Jorge Luiz de C. Henriques Junior², Nordeval C. Araújo²,
José Hermógenes R. Suassuna², Felipe da R. Henriques¹

¹ CEFET/RJ – Rio de Janeiro – RJ – Brasil

² HUPE – UERJ – Rio de Janeiro – RJ – Brasil

jefferson.anjos.1@aluno.cefet-rj.br,

{amaro.lima,gabriel.araujo,felipe.henriques}@cefet-rj.br,

{jorgehenriquesjr,nordevalaraujo}@gmail.com, rocco@uerj.br

Abstract. *The interpretation of renal ultrasound images depends on subjective visual inspection and anatomical kidney delineation. This paper proposes a computer vision pipeline that first segments the kidney and then analyzes suspicious hyperechoic regions near the renal pyramids. The models were retrained using the original kidneyUS images with CLAHE preprocessing. On a deduplicated test set of 70 images, U-Net achieved Dice of 0.9290, IoU of 0.8722, and precision of 0.9236, obtaining the highest Dice among the evaluated architectures. Applied to 4,479 external images, this model generated 3,534 candidate pseudomasks, which remain pending human review and are not included in the reported metrics.*

Resumo. *A interpretação de ultrassonografias renais depende de inspeção visual subjetiva e da delimitação anatômica do rim. Este artigo propõe um pipeline de visão computacional que primeiro segmenta o rim e depois analisa regiões hiperecogênicas suspeitas próximas às pirâmides renais. Os modelos foram retreinados com as imagens originais da base kidneyUS e aplicação de CLAHE no pré-processamento. Em um teste deduplicado de 70 imagens, a U-Net atingiu Dice de 0,9290, IoU de 0,8722 e precisão de 0,9236, obtendo o maior Dice entre as arquiteturas avaliadas. Aplicado a 4.479 imagens externas, esse modelo gerou 3.534 pseudomáscaras candidatas, ainda pendentes de revisão humana e ausentes das métricas apresentadas.*

1. Introdução

A ultrassonografia renal é acessível, não invasiva e amplamente utilizada na avaliação de alterações estruturais do rim. Na prática clínica, regiões muito claras no parênquima, especialmente próximas às pirâmides renais, podem indicar achados suspeitos que exigem análise cuidadosa.

A ecogenicidade cortical, especialmente quando comparada ao fígado, também é observada em pacientes com doença renal crônica. A razão entre a intensidade cortical renal e a intensidade de um órgão de referência pode reduzir parte da subjetividade da inspeção visual [Manley and O'Neill 2001].

Em paralelo, a literatura explora a segmentação renal e a extração de medidas quantitativas de intensidade e textura para investigar padrões associados a alterações renais [Chavhan et al. 2008]. Ainda assim, permanece uma lacuna entre segmentar o rim e examinar, dentro dele, pontos hiperecogênicos candidatos a achados suspeitos.

Esse recorte é importante porque a ultrassonografia apresenta ruído, sombras acústicas, variações de ganho e diferenças de enquadramento. Sem uma máscara

renal confiável, pontos muito claros podem ser confundidos com artefatos ou estruturas externas ao rim. Por isso, este trabalho parte da delimitação semântica do rim e, somente depois, transforma achados hiperecogênicos em medidas computacionais.

Nesse contexto, este artigo combina segmentação renal, expansão supervisionada por pseudomáscaras e extração de atributos intrarrenais. As imagens originais da base kidneyUS foram usadas para retrainar, com CLAHE (*Contrast Limited Adaptive Histogram Equalization*) [Zuiderveld 1994], quatro arquiteturas de segmentação da cápsula renal, entre as quais a U-Net apresentou o melhor *Dice* no teste deduplicado. A região renal delimitada alimenta uma segunda U-Net, responsável por identificar córtex, medula e complexo ecogênico central. As contribuições são: (i) consolidação do acervo com as imagens originais e suas anotações; (ii) geração rastreável de pseudomáscaras externas para revisão; (iii) comparação de segmentadores para delimitação renal; e (iv) segmentação das estruturas intrarrenais como suporte à análise exploratória de achados hiperecogênicos.

2. Contexto e Trabalhos Relacionados

A segmentação automática do rim em ultrassonografias constitui a primeira linha relacionada a este estudo. Arquiteturas como U-Net [Ronneberger et al. 2015], DeepLabV3 [Chen et al. 2018] e SegFormer [Xie et al. 2021] empregam diferentes estratégias para preservar bordas, capturar contexto multiescala e lidar com variações de textura. Neste trabalho, essas arquiteturas são comparadas na delimitação da cápsula renal, etapa que restringe as análises posteriores ao interior do rim.

A segunda linha envolve a quantificação da ecogenicidade renal relativa. Medidas entre a intensidade do córtex renal e a de um órgão de referência, como o fígado, podem aproximar critérios empregados por especialistas e reduzir parte da subjetividade da inspeção visual [Manley and O'Neill 2001].

A terceira linha compreende a extração de atributos quantitativos de intensidade e textura para investigar alterações renais [Chavhan et al. 2008]. As pirâmides renais estão localizadas na medula, entre o córtex e o sistema coletor, e constituem referências anatômicas internas na ultrassonografia. Alterações de ecogenicidade nessa região foram associadas à função renal em cenários clínicos específicos [Chavhan et al. 2008], motivando a análise de regiões hiperecogênicas próximas às pirâmides.

Além da delimitação externa do rim, este trabalho identifica córtex, medula e complexo ecogênico central. Essa representação não segmenta individualmente cada pirâmide renal, mas fornece uma localização anatômica mais específica para restringir a busca por achados hiperecogênicos.

A qualidade das segmentações é avaliada pela comparação entre a máscara produzida pelo modelo e a anotação manual. O coeficiente *Dice* [Dice 1945] mede a semelhança entre essas duas regiões, considerando duas vezes a área de interseção em relação à soma de suas áreas. A *Intersection over Union* (IoU), equivalente ao índice de Jaccard [Jaccard 1901], corresponde à razão entre a interseção e a união das regiões. Ambas variam de 0 a 1: valores próximos de 1 indicam maior concordância entre predição e referência. A precisão indica a proporção da região prevista que pertence à referência, enquanto o *recall* indica quanto da estrutura anotada foi recuperado.

Trabalhos também discutem a organização de dados de saúde para aprendizagem computacional [Lima et al. 2021] e o uso de aprendizado ativo para reduzir o custo de rotulação [Constantino et al. 2022]. Este estudo aproxima essas linhas em um *pipeline* no qual imagens e anotações são organizadas segundo sua proveniência, a cápsula delimita o

rim e as estruturas intrarrenais fornecem referências para a análise exploratória de regiões hiperecogênicas.

3. Metodologia

A metodologia organiza o processamento em três etapas encadeadas: localização da cápsula renal, verificação da segmentação pelo consenso entre modelos e identificação das estruturas intrarrenais. Essa organização delimita progressivamente a região analisada e reduz a influência de tecidos externos ao rim. A Figura 1 apresenta o fluxo geral.

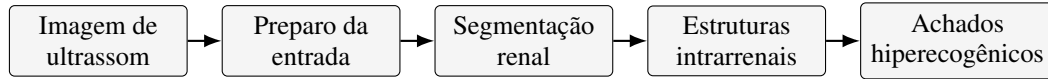


Figura 1. Pipeline metodológico: a cápsula delimita o rim, cujas estruturas internas orientam a análise de achados hiperecogênicos.

O estudo utiliza o *Open Kidney Ultrasound Data Set* (kidneyUS) [Singla et al. 2023], que contém anotações de dois observadores para cápsula, córtex, medula e complexo ecogênico central. A Tabela 1 resume os acervos utilizados e suas finalidades.

Tabela 1. Acervos auditados e finalidade no estudo.

Acervo	Reg.	Únicas	Finalidade
kidneyUS – cápsula	486	468	328 treino, 70 validação e 70 teste
kidneyUS – regiões internas	335	335	235 treino, 50 validação e 50 teste
MONAI	4.479	4.479	3.534 pseudomáscaras para revisão

Para a cápsula renal, foram usados 486 registros do primeiro anotador, correspondentes a 468 imagens únicas. Após deduplicação e agrupamento por exame, obtiveram-se 328 imagens de treino, 70 de validação e 70 de teste. Para a tarefa intrarrenal, 335 imagens foram divididas por paciente em 235, 50 e 50 imagens. O MONAI foi utilizado apenas para geração de pseudomáscaras, sem participação nas métricas manuais.

Na primeira etapa, U-Net, UNet++, DeepLabV3-ResNet50 e SegFormer-B0 foram treinados para segmentar a cápsula, com CLAHE, aumento de dados apenas no treino, AdamW, taxa inicial de 10^{-4} e perda BCE+Dice. A U-Net, selecionada pelo maior *Dice*, foi aplicada ao MONAI. A inferência combinou a imagem original e seu espelhamento horizontal; após binarização pelo limiar 0,35, manteve-se o maior componente conectado e a máscara foi restaurada à resolução original. Dessa forma as pseudomáscaras foram ordenadas pela confiança média dos pixels positivos, definida por:

$$C = \frac{1}{|M|} \sum_{i \in M} p_i, \quad (1)$$

em que M é a pseudomáscara e p_i a probabilidade prevista para o pixel i . Esse valor auxilia a priorização da revisão, mas não valida automaticamente a máscara [Lee 2013]. Para controlar as pseudomáscaras externas, as mesmas imagens foram processadas pela DeepLabV3-ResNet50. A concordância entre modelos foi estimada pelo coeficiente *Dice* [Dice 1945] entre a máscara da U-Net, M_U , e a máscara da DeepLabV3, M_D , da seguinte forma:

$$S(M_U, M_D) = \frac{2|M_U \cap M_D|}{|M_U| + |M_D|}. \quad (2)$$

O score da Equação (2) foi usado como medida complementar de consenso entre arquiteturas. Os limiares foram calibrados por validação cruzada em cinco

folds agrupados por exame, de modo que cada uma das 468 imagens manuais fosse avaliada apenas por modelos que não a utilizaram no treino. Os percentis 10 e 25, arredondados, definiram concordância baixa ($< 0,89$), intermediária ($0,89$ a $0,94$) e alta ($\geq 0,94$). Predições isoladas também receberam prioridade de revisão. A Figura 2 mostra as máscaras individuais e a sobreposição dos contornos da U-Net, em verde, e da DeepLabV3, em vermelho.

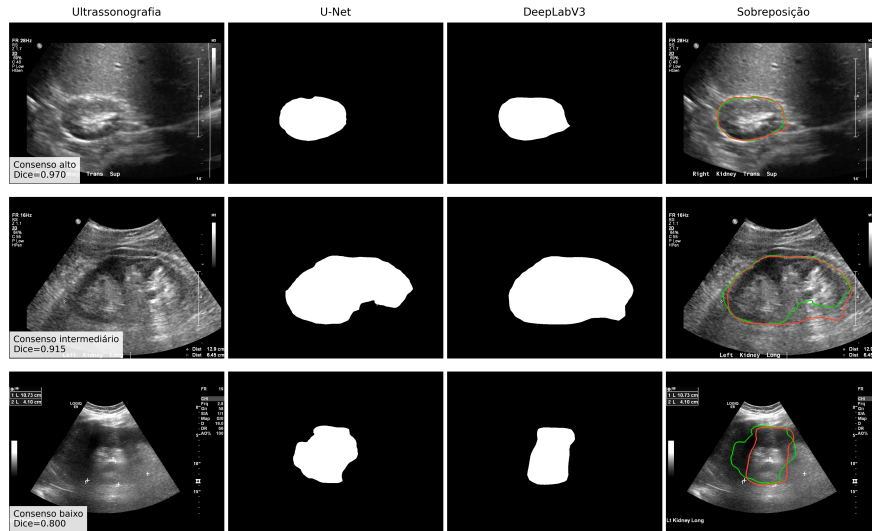


Figura 2. Consenso entre U-Net (verde) e DeepLabV3 (vermelho): concordância alta, intermediária e baixa.

O consenso não gera uma terceira máscara: ele verifica se arquiteturas distintas localizaram uma região semelhante e organiza a revisão das pseudomáscaras externas. Na cascata proposta, uma cápsula considerada adequada forma a região de interesse (ROI) da terceira etapa.

Para a segmentação intrarrenal, U-Net e DeepLabV3-ResNet50 foram retreinadas na mesma divisão de 235 imagens de treino, 50 de validação e 50 de teste. Ambas recebem três canais — ROI em tons de cinza, ROI com a região externa zerada e máscara da cápsula — e predizem córtex, medula e complexo ecogênico central, com CLAHE, AdamW, taxa de 10^{-4} e perda entropia cruzada+*Dice*. A U-Net constitui o modelo da cascata e a DeepLabV3, a referência comparativa.

A Figura 3 exemplifica a saída da etapa intrarrenal.



Figura 3. Encadeamento da etapa intrarrenal: imagem original, anotação manual e predição multiclasse da U-Net.

A Figura 3 apresenta, da esquerda para a direita, a ultrassonografia, a referência manual e a predição da U-Net multiclasse. Ciano representa o córtex, amarelo a medula e laranja o complexo ecogênico central. O exemplo mostra que, após a delimitação da cápsula, as classes ficam restritas à região renal e reproduzem sua organização anatômica geral, embora existam diferenças nos limites de cada compartimento. A classe

Medulla segue a anotação do *kidneyUS* e não representa pirâmides individualmente segmentadas.

4. Resultados

A primeira etapa avaliou a segmentação da cápsula renal. Conforme a Tabela 2, nas 70 imagens do teste deduplicado, a U-Net apresentou os melhores valores de *Dice* (0,9290), *IoU* (0,8722), precisão (0,9236) e *recall* (0,9440), sendo selecionada para localizar o rim na cascata. O SegFormer-B0, embora ligeiramente inferior nas métricas de sobreposição, obteve a maior velocidade de inferência (FPS - *Frames Per Second* - em valores médios).

Tabela 2. Segmentação da cápsula no teste deduplicado.

Modelo	Dice	IoU	Precisão	Recall	FPS
U-Net	0,9290	0,8722	0,9236	0,9440	18,11
SegFormer-B0	0,9223	0,8598	0,9117	0,9426	31,34
DeepLabV3 R50	0,9203	0,8606	0,9219	0,9338	22,26
UNet++	0,9099	0,8419	0,8910	0,9419	16,57

O controle por consenso foi calibrado nas 468 imagens manuais por validação cruzada em cinco *folds*. O consenso médio foi 0,9416, com mediana de 0,9627, e apresentou correlação (*r*) de 0,840 com o menor *Dice* frente à anotação manual. Os *Dices* médios fora da amostra foram 0,9150 para a U-Net e 0,9243 para a DeepLabV3. Assim, o consenso foi usado como indicador operacional de qualidade e priorização, não como substituto da validação manual.

Nas 4.479 imagens externas, 1.529 (34,1%) apresentaram concordância alta entre os modelos, correspondente a *Dice* de consenso igual ou superior a 94%, e 699 (15,6%) ficaram na faixa intermediária, entre 89% e 94%. A revisão prioritária reuniu 1.741 imagens (38,9%): 1.204 com concordância inferior a 89% e 537 segmentadas por apenas um modelo, das quais 435 somente pela DeepLabV3 e 102 somente pela U-Net. Em 510 imagens (11,4%), nenhum modelo delimitou o rim. Assim, a U-Net gerou 3.534 pseudomáscaras entre as 4.479 imagens processadas. Como o subconjunto externo não possui anotação manual neste estudo, essas porcentagens expressam concordância e cobertura da triagem, e não probabilidade de acerto ou acurácia externa.

Por fim, U-Net e DeepLabV3 foram comparadas na segmentação intrarrenal usando as mesmas 50 imagens de teste com referência manual. De acordo com a Tabela 3, a U-Net obteve os maiores valores de *Dice* para córtex (0,7075) e medula (0,7237), enquanto a DeepLabV3 foi ligeiramente superior no complexo ecogênico central - CEC (0,8497). Na média das três classes, a U-Net alcançou *Dice* de 0,7594 e *IoU* de 0,6163, superando os valores de 0,7492 e 0,6045 da DeepLabV3.

Tabela 3. Comparação intrarrenal no teste manual de 50 imagens.

Modelo	Córtex	Medula	CEC	Dice médio	IoU médio
U-Net	0,7075	0,7237	0,8469	0,7594	0,6163
DeepLabV3 R50	0,6822	0,7156	0,8497	0,7492	0,6045

Embora a diferença média seja pequena, a U-Net foi mantida como modelo intrarrenal da cascata por apresentar o melhor resultado agregado e maior *Dice* em duas das três estruturas. O melhor desempenho de ambas as arquiteturas ocorreu no complexo ecogênico central, enquanto o córtex permaneceu como a classe mais difícil de delimitar.

5. Discussão

A qualidade das imagens foi um dos fatores observados nas falhas de segmentação da cápsula. O baixo contraste, sombras acústicas, ruído e perda de detalhes dificultam a

identificação dos limites renais. A Figura 4 apresenta um contraste qualitativo entre dois exemplos externos.

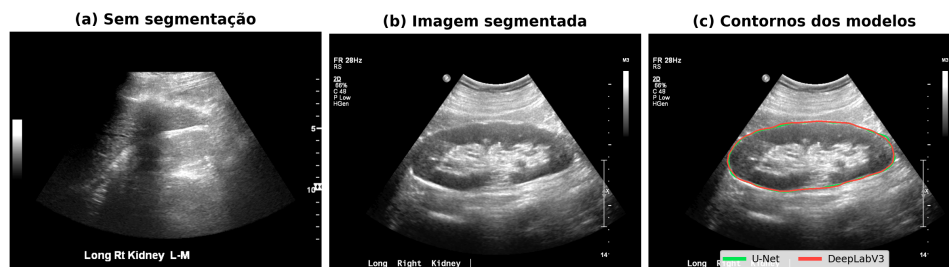


Figura 4. Contraste qualitativo no conjunto externo: (a) ausência de predição; (b) imagem com limites renais mais nítidos; e (c) contornos da U-Net (verde) e da DeepLabV3 (vermelho), com consenso *Dice* de 0,9888.

No primeiro caso, a baixa diferenciação visual da cápsula coincide com a ausência de predição pelos dois modelos; no segundo, os limites são mais evidentes e as segmentações apresentam elevada concordância. Sem referência manual para essas imagens, o exemplo não demonstra isoladamente uma relação causal.

A cascata permanece dependente da primeira segmentação: uma máscara incompleta ou com tecido adjacente altera a ROI e pode comprometer a identificação das estruturas internas. As métricas intrarrenais também devem ser interpretadas considerando o tamanho do teste e a variabilidade entre anotadores.

A pseudorrotulação MONAI constitui engenharia de dados, não resultado clínico. O consenso prioriza a revisão, mas somente uma referência manual confirma a localização correta do rim. Experimentos futuros devem incorporar apenas máscaras revisadas, manter validação e teste manuais invariantes e usar referência clínica ou histológica para investigar fibrose.

6. Conclusão

Este trabalho apresentou um *pipeline* rastreável para análise de ultrassonografias renais em três etapas encadeadas: segmentação da cápsula renal, verificação por consenso entre modelos e segmentação das estruturas intrarrenais. A auditoria por *hash* do acervo *kidneyUS* identificou 18 registros redundantes distribuídos em nove grupos, cinco deles atravessando partições originais, motivando a reorganização das divisões por exame antes do treinamento.

Na delimitação da cápsula, a U-Net obteve *Dice* de 0,9290 e IoU de 0,8722 no teste deduplicado de 70 imagens, superando as demais arquiteturas avaliadas. O consenso entre U-Net e DeepLabV3 ($r = 0,840$ com o *Dice* individual) foi adotado como critério operacional para priorizar a revisão humana das 3.534 pseudomáscaras geradas sobre as 4.479 imagens externas, sem substituir a validação por especialistas. Na etapa intrarrenal, a U-Net alcançou *Dice* médio de 0,7594 nas 50 imagens de teste com referência manual.

As diferenças entre arquiteturas na etapa intrarrenal não foram submetidas a testes de significância estatística, e a avaliação de ponta-a-ponta do *pipeline* permanece como trabalho futuro. Além disso, pretende-se incorporar apenas pseudomáscaras revisadas por especialistas e estabelecer referência clínica ou histológica para investigar, de forma validada, a associação entre regiões hiperecogênicas e alterações funcionais do rim.

Referências

Chavhan, G. B., Parra, D. A., Oudjhane, K., et al. (2008). Renal pyramid echogenicity in ureteropelvic junction obstruction: correlation between altered echogenicity and differential renal function. *Pediatric Radiology*, 38(10):1063–1069.

- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *European Conference on Computer Vision (ECCV)*, pages 801–818.
- Constantino, K., Cruz, V. A. L., Zucheratto, O. M. M., França, C., Carvalho, M., Silva, T. H. P., Laender, A. H. F., and Gonçalves, M. A. (2022). Segmentação e classificação semântica de trechos de diários oficiais usando aprendizado ativo. In *Anais do XXXVII Simpósio Brasileiro de Banco de Dados (SBB D)*, pages 304–316. Sociedade Brasileira de Computação.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:547–579.
- Lee, D.-H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on Challenges in Representation Learning, ICML*, volume 3.
- Lima, D. M., Moreno, R. A., Pires, F. A., and Gutierrez, M. A. (2021). Uma proposta de data lake para pesquisa em saúde a partir de data pools multicêntricos interoperáveis. In *Anais do XXXVI Simpósio Brasileiro de Banco de Dados (SBB D)*, pages 367–372. Sociedade Brasileira de Computação.
- Manley, J. A. and O’Neill, W. C. (2001). How echogenic is echogenic? quantitative acoustics of the renal cortex. *American Journal of Kidney Diseases*, 37(4):706–711.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241.
- Singla, R., Ringstrom, C., Hu, G., Lessoway, V., Reid, J., Ngan, C., and Rohling, R. (2023). The open kidney ultrasound data set. In *Simplifying Medical Ultrasound*, volume 14337 of *Lecture Notes in Computer Science*, pages 155–164. Springer Nature Switzerland.
- Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090.
- Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In Heckbert, P. S., editor, *Graphics Gems IV*, pages 474–485. Academic Press.