Souza *et al. BMC Nephrology* (2025) 26:544
https://doi.org/10.1186/s12882-025-04469-6

BMC Nephrology

RESEARCH

Open Access

The problem of segmenting global glomerulosclerosis in gigapixel histopathological images: the borderless glomeruli

Luiz Souza[1,3], Jefferson Silva[1,2], Marcelo Mendonça[1,4], José Nathan[6], Angelo Duarte[5], Pinaki Sarder[7], Washington L. C. dos-Santos[6*] and Luciano Oliveira[1,6*]

Abstract

Background Accurately segmenting glomeruli in kidney whole slide images (WSIs) is essential for advancing automation in renal pathology but remains challenging in cases of global glomerulosclerosis, where Bowman's capsule boundaries are often unclear. Conventional machine learning (ML) models perform well on normal glomeruli but struggle with sclerotic cases due to the lack of distinct structural cues. This study investigates the use of the foundation model segmentation generative pre-trained transformer (SegGPT) to address this limitation.

Methods We conducted experiments at both the patch and WSI levels on a private dataset to evaluate the performance of SegGPT foundation model against three non-foundation architectures, U-Net, U-Net3+, and SwinTransformer + U-Net, trained with and without fine-tuning.

Results The study revealed high segmentation performance for normal glomeruli, with non-foundation models achieving mean Dice similarity coefficient (mDice) scores of up to 0.94. For segmental sclerosis, performance was moderate, with scores reaching up to 0.73. In contrast, the segmentation of globally sclerotic glomeruli proved substantially more challenging: Models trained only on normal samples yielded mDice scores below 0.03, and even with fine-tuning on mixed datasets, WSI-level performance remained limited (mDice < 0.16). With only few annotated examples, SegGPT demonstrated markedly superior performance in this scenario, achieving up to 0.43 at the WSI level and 0.74 at the patch level. However, its performance under idealized conditions also reveals limitations in clinical generalization.

Conclusion While conventional models perform well on normal and segmentally sclerotic glomeruli, their performance declines sharply in globally sclerotic cases, even with fine-tuning. SegGPT showed better generalization in these challenging scenarios, particularly at the patch level. However, its limited performance at the WSI level underscores the difficulty of translating patch-level accuracy to full-slide inference, where contextual ambiguity is

*Correspondence:
Washington L. C. dos-Santos
washington.santos@fiocruz.br
Luciano Oliveira
lrebouca@ufba.br

Full list of author information is available at the end of the article

greater. These results expose a persistent gap between controlled experimental setups and real-world conditions, reinforcing the need for more realistic evaluation protocols to advance clinical applicability.

Keywords Automatic segmentation, Glomerulus, Sclerosis

Background

The structural evaluation of renal biopsies increasingly relies on digital pathology, where gigapixel WSIs enable detailed navigation and interpretation of tissue compartments [1–3]. Among these, glomeruli are key diagnostic units — central to the assessment of both active and chronic kidney injury. In particular, chronic lesions such as glomerulosclerosis introduce notable variation in glomerular morphology, which can directly impact diagnostic accuracy and consistency. Global glomerulosclerosis reflects the end stage of processes related to aging or renal injury and primarily indicates chronicity. Segmental sclerosis also signals chronicity but serves as a diagnostic or classificatory marker in conditions such as focal segmental glomerulosclerosis (FSGS) and IgA nephropathy. While global glomerulosclerosis may simply represent age-related change, segmental sclerosis more often indicates disease. Given its role as a chronicity marker, quantifying globally sclerotic glomeruli is a key first step in renal biopsy assessment.

Given the diagnostic weight placed on glomerular assessment, particularly in diseases with heterogeneous glomerular involvement, several studies have proposed segmentation frameworks based on ML and deep neural networks [4–18]. However, most models are trained and validated on datasets dominated by normal glomeruli and do not explicitly stratify performance by glomerular subtype [19–28]. As a result, their reported accuracy may overestimate clinical utility, particularly in cases of advanced sclerosis or mixed lesion distribution.

To contextualize the position of our study within the field, Table 1 summarizes relevant works according to three inclusion criteria: (i) perform glomerular segmentation, (ii) use WSI-based data, and (iii) are indexed in PubMed. Among the works reviewed, U-Net and its variants emerge as the prevailing architectures; staining protocols most commonly include hematoxylin and eosin (HE), periodic acid-Schiff (PAS), and periodic acid-methenamine silver (PAMS); and only a minority of studies address the two glomerular sclerosis categories: segmental and global.

The scarcity of public datasets, the low representation of sclerotic glomeruli, and the high cost of annotation continue to constrain the development and evaluation of segmentation models, particularly for lesions that diverge from canonical glomerular morphology. Even among studies that include sclerotic glomeruli [4–8, 10, 11], class-specific performance is often underreported, raising concerns about the applicability of these models to real-world, heterogeneous renal pathology.

Quantitative analyses focusing on global glomerulosclerosis have been explored in relatively few studies [9, 12, 13]. Jiang et al. [9] analyzed glomeruli with global sclerosis in comparison to normal glomeruli and those exhibiting other pathological lesions, such as moderate to severe mesangial hypercellularity or expansion, crescents, and apparent endothelial proliferation. The authors used training and test datasets within the same glomerulus class domain, employing the mean Intersection over Union (mIoU) metric to evaluate the three classes. Similarly, Yu et al. [12] developed a framework for segmenting glomerular structures (Bowman's capsule, tuft, and mesangium), two cellular components (mesangial cells and podocytes), and nine lesion classes (e.g., adhesion, capsular drop, global sclerosis, hyalinosis, mesangial lysis, microaneurysm, nodular sclerosis, mesangial expansion, and segmental sclerosis) across human and rodent WSIs. Wang et al. [13] proposed Glo-Net, a dual-task deep learning (DL) network for simultaneous glomerular segmentation and classification in renal pathology. Their framework was evaluated using 253 PAS-stained WSIs from three independent datasets (Gulou Hospital, Zhongda Hospital, and the AIDPATH Project), encompassing ten histological classes, including crescent, mesangial hyperplasia, normal glomeruli, globally sclerotic glomeruli, segmentally sclerotic glomeruli, cellular crescents, fibrous crescents, glomerular sclerosis, incomplete glomeruli, and other glomerular types.

Those studies have reported strong results in glomerular segmentation; however, important limitations remain, particularly in their handling of globally sclerotic glomeruli. For instance, Jiang et al. [9] adopted a preprocessing strategy in which WSIs were cropped into 2048×2048-pixel patches, followed by manual verification to exclude irrelevant regions and ensure that each glomerulus was fully contained within a patch. This level of manual curation likely impacted segmentation performance — a point that will be further explored in our analysis. Similarly, Yu et al. [12] used 512×512-pixel patches derived from high-magnification scans; although this work do not explicitly state whether globally sclerosed glomeruli were manually isolated, their results suggest such an approach, as segmentation performance on globally sclerosed glomeruli was unexpectedly higher than on segmentally sclerosed ones. Wang et al. [13], on the other hand, observed the expected performance decrease when segmenting globally sclerosed glomeruli compared to

Table 1 Summary of related work for glomerulus segmentation. Dashes indicate instances where information was unreported

Year	Reference	Method	#WSI	#Glomerulus	Stain	Classes	Dataset
2018	[4]	FCN, M-FCN, U-Net	15	944	PAS	Normal, Sclerosis	Private
2018	[5]	CNN(VGG-16), LoG, Blob-detection, Linear Regression	48	3867	HE	Normal, Sclerosis	Private
2019	[19]	SW-CNN, U-Net-S, U-Net-D	24	-	PAS	-	Private
2019	[6]	U-Net	60	238	PAS	Normal, Sclerosis	Private
2020	[7]	U-Net, SegNet + AlexNet	47	1245	PAS	Normal, Sclerosis	Private
2020	[8]	SegNet, DeepLabv3+	26	2772	PAS	Normal, Sclerosis	Private
2021	[20]	Mask-RCNN, U-Net, DeepLabv3	61	1334	HE, PAS, PAMS	-	Private
2021	[9]	Cascade Mask-RCNN	348	8665	PAS, PAMS, TRI	Normal, Global Sclerosis, Other Lesions	Private
2021	[21]	U-Net	459	1196	HE, PAS, PAMS, TRI	-	Private
2021	[22]	Modified U-Net	168	2611	PAS	-	Private
2021	[10]	U-Net	258	24,133	HE	Normal, Sclerosis	Public
2021	[11]	U-Net	51	2429	HE, PAS	Normal, Sclerosis	Public
2022	[23]	DS-FNet	665	5309	HE, PAS, PAMS, TRI	-	Public Private
2023	[24]	U-Net, U-Net3+, Res-U-Net DeepLabV3+, MA-Net	60	1430	HE, PAS PAMS	-	Public Private
2023	[25]	Cascade R-CNN	875	23,477	PAS	-	Private
2023	[26]	Omni-Seg (Residual-U-Net + DoDNet)	459	1,196	HE PAS PAMS TRI	-	Private
2024	[27]	PrPSeg (Residual-U-Net)	459	1,196	HE PAS PAMS TRI	-	Private
2024	[28]	HATs (EfficientSAM)	459	1,196	HE PAS PAMS TRI	-	Private
2024	[12]	Glo-In-One-v2 (Residual-U-Net)	368	23,529	-	Normal, Segmental Sclerosis, Global Sclerosis, Other Lesions,	Private
2025	[13]	Glo-net (SwinTransformer + U-Net++)	253	4,966	PAS	Normal, Segmental Sclerosis, Global Sclerosis, Other Lesions	Private
2025	**Ours**	**U-Net**, **U-Net3+**, **SwinTransformer + U-Net**, **Fine-tuned versions**, **SegGPT**	**83**	**800**	**HE**, **PAS**, **PAMS**	**Normal**, **Segmental Sclerosis**, **Global Sclerosis**	**Private**

normal and segmentally sclerosed glomeruli. Nevertheless, their reported performance metrics across all cases are remarkably high (mIoU > 0.8 for globally sclerosed glomeruli and mIoU > 0.9 for normal and segmentally sclerosed glomeruli), which seems inconsistent with the inherent difficulty of segmenting heavily degraded structures. This discrepancy raises questions, especially given the lack of detailed information about their methodology. Without clarity regarding whether their results relied on human-assisted preprocessing, it is difficult to fully assess the validity of their findings.

Accurately segmenting globally sclerotic glomeruli directly within WSIs remains a significant challenge, underscoring the need for continued research in more realistic, whole-slide contexts. Although post-processing approaches [29] have been proposed to enhance existing segmentation methods, they cannot overcome severe failures when a glomerulus is not detected at all — a situation frequently observed in cases of global sclerosis.

Ideally, segmentation should be performed directly on WSIs, where glomeruli must be distinguished from surrounding renal structures such as tubules, interstitium, and vessels. In contrast, segmentation on cropped images tends to be simpler, as the glomerulus is centrally located, visually prominent, and often isolated from confounding tissue components. Approaches that begin at the WSI level are better aligned with real-world clinical workflows and offer more effective support for pathologists' decision-making.

In our study, we adopted a less common but clinically motivated direction: We performed segmentation starting from full WSIs and only later assessed performance on automatically cropped glomeruli. Both the WSI- and patch-level segmentations were evaluated using the mDice. While Dice is known to be more sensitive to small structures, emphasizing true positives and penalizing false negatives, IoU tends to be more influenced by false positives. We prioritized mDice to better reflect performance in the context of sclerotic glomeruli, which often occupy a small and morphologically ambiguous area. The main contribution of this paper is then to elucidate the ill-posed problem of automatic glomerulus segmentation, revealing critical aspects necessary for effectively paving the way to solve the segmentation problem involving few examples, difficult-to-label image objects, and highly unbalanced classes (with a predominance of glomeruli with well-defined boundaries). To this end, we conducted experiments on a dataset of histological images from representative diagnostic cases, encompassing a variety of glomeruli, both normal and with sclerotic lesions.

Methods

A clinical-practice-driven approach for sclerotic glomerulus segmentation

Figure 1 presents examples of glomeruli across a spectrum of sclerotic involvement: (a) normal glomerulus, (b) segmentally sclerotic glomerulus, and (c) globally sclerotic glomerulus. While normal and segmentally affected glomeruli often retain Bowman's space and identifiable nuclear detail, globally sclerotic glomeruli exhibit extensive matrix accumulation, tuft obliteration, and poorly defined borders, blending into the surrounding interstitium [30]. Such morphological degradation challenges both manual review and algorithmic segmentation, particularly when the structural cues used for delineation are lost. Although existing computational approaches perform well in segmenting normal glomeruli, they often fail to generalize to sclerotic phenotypes, especially in globally sclerotic cases. This gap is clinically relevant, as reliable glomerular segmentation underpins downstream analyses such as lesion quantification and pattern classification in renal pathology.

The first assumption of our study is that segmentation must be conducted under similar conditions to pathology practice, without any manual intervention. Since WSIs of kidney biopsies are typically gigapixel-sized, the WSIs are first automatically divided into regular tiles (or patches). These tiles preserve patient-WSI associations, with all patches being generated automatically from the original WSIs.

The proposed framework is illustrated in Fig. 2 and comprises training and prediction phases. Since this framework is agnostic regarding the segmentation network and based on the literature review, we employed four powerful methods: Vanilla U-Net [31], U-Net 3+ [32], SwinTransformer [33] as the backbone of a U-Net, and SegGPT [34]. The first three models are traditional supervised DL methods, whereas SegGPT follows an N-shot learning paradigm, using one or a few reference query images to guide the segmentation process.

The U-Net is a widely adopted encoder–decoder architecture in medical image segmentation, characterized by symmetric skip connections between each encoder and decoder layer, forming a U-shaped structure.

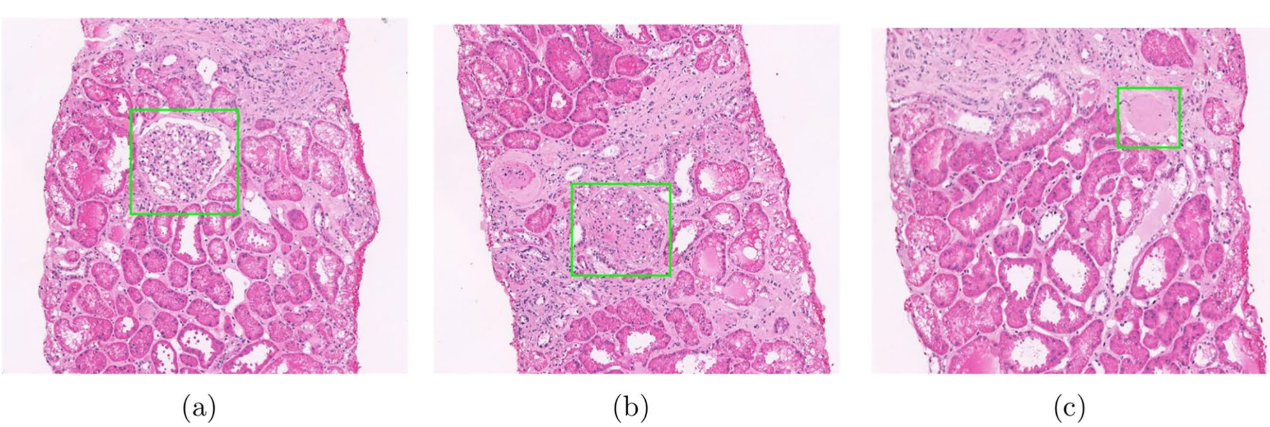


Fig. 1 Examples of glomeruli (green bounding box) from the WSI data sets. (**a**) Normal glomerulus – A healthy glomerulus with a well-preserved capillary tuft, clearly defined Bowman's space, and Bowman's capsule with defined borders. (**b**) Segmentally sclerotic glomerulus – A glomerulus showing partial sclerosis, with less than 100% of the tuft affected. (**c**) Globally sclerotic glomerulus – A glomerulus with complete (100%) sclerosis of the tuft, confirmed across serial sections

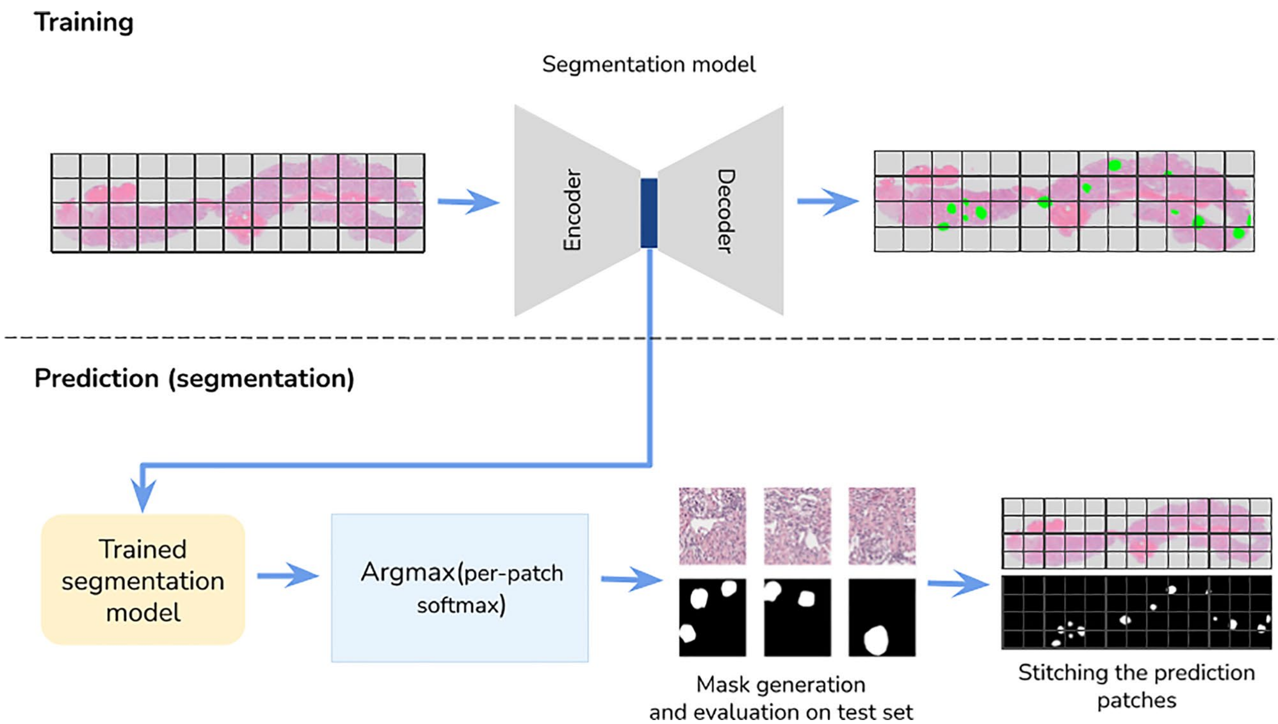


Fig. 2 Proposed framework for ensuring comprehensive glomerulus segmentation. Glomeruli are manually annotated by pathologists from human kidney biopsy sections and automatically segmented into patches - these patches serve as inputs for training segmentation networks (agnostically defined here); the best model from each network is selected to predict the pixel class of the glomeruli in each patch; finally, the resulting masks are generated, and evaluated using a Dice metric

Several advanced variants of U-Net have been proposed to enhance its performance. U-Net 3+, for instance, redesigns skip connections to aggregate multi-scale features more effectively by fusing encoder and decoder feature maps across multiple scales, allowing accurate segmentation with fewer parameters. Another network is the SwinTransformer, which integrates the SwinTransformer blocks into the encoder, bottleneck, and decoder stages. This hybrid model leverages the hierarchical self-attention mechanisms of transformers to capture long-range dependencies while preserving the advantages of the U-Net structure through skip connections. Finally, SegGPT is a generalist model developed for context-aware segmentation tasks; it is based on a vision transformer architecture and leverages a foundation model approach. SegGPT leverages feature representations from a few query image samples during inference.

To ensure a consistent and realistic evaluation across all models, we adopted a full-slide inference framework in which each segmenter processes WSIs directly, without the need for manual cropping into patches (see Fig. 2). Instead, the WSIs are automatically divided into overlapping tiles, and segmentation predictions are made on each tile. A stitching algorithm is then applied to seamlessly merge the tile-level predictions into a coherent whole-slide segmentation mask. This approach allows for direct comparison of the models under conditions that reflect real-world clinical workflows. To support this evaluation, a specialized dataset of annotated WSIs was curated by our pathology team, as detailed in the following section.

Datasets

The dataset[1] used in this paper originates from two newly introduced sets named Normal and Sclerosis. The 'Normal' dataset extends the dataset found in [24], now featuring 45 WSIs (5 patients) of human kidney biopsies. The 'Sclerosis' dataset is newly created and includes 37 human kidney Sect. (33 patients). The 'Mixed' dataset is also new and contains 1 human kidney Sect. (1 patient). The renal biopsies used to build the dataset were fixed in formalin-acetic acid-alcohol to preserve their histological structure, later included in paraffin. Images were finally captured using a VS 110 Olympus scanner with 40× magnification for the normal dataset, and a Zeiss Imager Z2 scanner with 20× magnification for the sclerosis and mixed datasets. The main characteristics of these datasets are summarized in Table 2.

Normal dataset: It comprises 45 high-resolution WSIs obtained from 5 patients, with tissue Sect. (2 μm thick) stained using HE, PAS, and PAMS: 15 images per stain.

[1] To support reproducibility, the dataset will be made available for research purposes upon request.

Table 2 Summary of the characteristics of the datasets

Dataset	Stain	# WSI		# Glomerulus		# Glomerulus by class			Average Pixels
		Train	Test	Train	Test	Normal	Segmental	Global	(height x width)
Normal	HE	12	3	194	27	221	-	-	23,203 × 10,639
	PAS	12	3	149	61	210	-	-	22,795 × 11,075
	PAMS	12	3	133	77	210	-	-	22,698 × 11,825
	ALL	36	9	476	165	641	-	-	22,899 × 11,180
Sclerosis	HE	-	19	-	71	-	48	23	29,211 × 15,818
	PAS	-	18	-	76	-	56	20	26,453 × 14,193
	ALL	-	37	-	147	-	104	43	27,869 × 15,028

All glomeruli in these slides exhibit normal morphology with no visible lesions. For model development, 36 WSIs from 4 patients (12 per stain) were used for training, and 9 WSIs from the remaining patient (3 per stain) were reserved for testing. This split preserves balanced stain distribution across both sets and reflects the entirety of the available data (3 sections per patient × 5 patients).

Sclerosis dataset: The dataset consists of 37 high-resolution WSIs from 33 patients diagnosed with FSGS, using 2 µm-thick sections stained with either HE or PAS. The glomeruli in these slides exhibit segmental or global glomerulosclerosis. For model testing, all available slides were used: 19 HE-stained WSIs (from 15 HE-only patients and 4 patients with both stains) and 18 PAS-stained WSIs (from 14 PAS-only patients and the same 4 with both stains), totaling 37 WSIs.

Mixed dataset: The dataset includes a single high-resolution WSI from one patient, stained with HE on 2 µm-thick tissue sections. Glomeruli within this slide are categorized into three classes: normal, segmental sclerosis, and global sclerosis. This WSI was used exclusively for training via fine-tuning and for class calibration in the SegGPT model.

For per-crop evaluation, 12 glomerular patches were manually extracted from the mixed dataset, comprising 4 patches per class.

Implementation details

U-Net, U-Net 3+, SwinTransformer + U-Net, and SegGPT methods were implemented using the PyTorch framework [35] version 1.9.1. The initial setup involved loading all networks with weights pre-trained on the Human BioMolecular Atlas Program (HuBMAP) dataset [36]. The network models were trained across 30 epochs with a batch size of 8, a warm-up learning rate scheduler at a maximum of 0.0001, and a weight decay of 0.00001, using a loss that combines binary cross-entropy (BCE) [37] and Lovasz [38] loss functions. During network training, the best-performing epoch was selected based on the highest Dice score and the lowest combined loss value. The WSIs were automatically tiled into patches of 1,024 × 1,024 pixels, with a padding size of 256 pixels. To manage the high memory footprint and enhance training speed, each patch was resized to 320 × 320 pixels before being used for training. All experiments were run on a laptop with AMD Ryzen 5 3600, 32GB RAM, and a GeForce RTX 3080 NVIDIA GPU containing 10GB of memory. The pre-trained weights were fine-tuned on the top layers, aiming to keep the rich features previously learned. To increase input variability during training, we applied the following online data augmentation techniques: resizing, horizontal and vertical flipping, random 90-degree rotations, shift-scale-rotate, Gaussian noise, Gaussian blur, random brightness/contrast adjustments, and random changes in hue and saturation.

Results

Given that normal glomeruli (see Fig. 1a) represent the majority in our dataset, we aim to demonstrate that this type of canonical glomerulus can aid in the segmentation of glomeruli affected by segmental glomerulosclerosis (see Fig. 1b). However, this convenience does not necessarily extend to glomeruli affected by global sclerosis (see Fig. 1c), even when a fine-tuning strategy is applied to samples of globally sclerotic glomeruli.

The primary goal of this study is to work within the framework depicted in Fig. 2. Additionally, we performed a per-crop evaluation following [9], in which the models were tested on individual glomerulus patches extracted from WSIs. This complementary analysis serves two purposes: First, it isolates the segmentation task from the challenges of glomerulus localization within the complex WSI background, allowing us to assess the intrinsic ability of each model to delineate glomerular boundaries; second, it provides an upper bound on segmentation performance under idealized conditions, where the target structure is fully contained within the input and free from distracting contextual features such as interstitium, vessels, or staining artifacts. By contrasting per-crop and WSI-level results, we can better understand the gap between model capability in controlled settings and their robustness in realistic diagnostic scenarios.

Considering the splits defined in Tables 2, 3, 4 and 5, and 6 summarize the results found by our proposed study. The metric used to assess the performance was mDice. The Dice score is calculated as twice the area of

Table 3 Mean ± standard deviation of dice scores evaluated on the normal WSI test set. Results are presented for three segmentation networks trained on the normal training set, as well as for the SegGPT model using 6 (2 per class: normal, segmental, Global) and 12 (4 per class) query examples

Stain	U-Net	U-Net3+	SwinTransformer + U-Net	SegGPT#6	SegGPT#12
HE	0.923 (± 0.014)	0.864 (± 0.022)	**0.935 (± 0.008)**	0.891 (± 0.109)	0.804 (± 0.299)
PAS	0.915 (± 0.005)	**0.931 (± 0.022)**	0.907 (± 0.009)	0.728 (± 0.317)	0.711 (± 0.292)
PAMS	**0.937 (± 0.006)**	0.926 (± 0.006)	0.850 (± 0.048)	0.776 (± 0.272)	0.805 (± 0.194)
ALL	**0.919 (± 0.018)**	0.908 (± 0.028)	0.918 (± 0.033)	0.798 (± 0.251)	0.774 (± 0.261)

Table 4 Mean ± standard deviation of dice scores evaluated on the segmental WSI test set. Results are presented for three segmentation networks trained on the normal training set, as well as for the SegGPT model using 6 (2 per class: normal, segmental and Global) and 12 (4 per class) query examples

Stain	U-Net	U-Net3+	SwinTransformer + U-Net	SegGPT#6	SegGPT#12
HE	**0.720 (± 0.192)**	0.575 (± 0.131)	0.650 (± 0.150)	0.508 (± 0.448)	0.646 (± 0.343)
PAS	**0.730 (± 0.197)**	0.643 (± 0.184)	0.702 (± 0.156)	0.548 (± 0.355)	0.593 (± 0.335)
ALL	**0.728 (± 0.163)**	0.667 (± 0.163)	0.675 (± 0.186)	0.529 (± 0.310)	0.618 (± 0.336)

Table 5 Mean ± standard deviation of dice scores for globally sclerotic glomerulus segmentation, evaluated on the global sclerosis WSI test set. Results are shown for three segmentation networks trained on the normal training set, without fine-tuning

Input	Stain	U-Net	U-Net3+	SwinTransformer + U-Net
WSI	HE	0.020 (± 0.018)	0.000 (± 0.000)	0.000 (± 0.000)
	PAS	0.002 (± 0.001)	0.027 (± 0.022)	0.017 (± 0.011)
	ALL	0.005 (± 0.002)	0.020 (± 0.013)	0.004 (± 0.002)
Per-crop	HE	0.000 (± 0.000)	0.000 (± 0.000)	0.035 (± 0.164)
	PAS	0.087 (± 0.268)	0.225 (± 0.282)	0.002 (± 0.010)
	ALL	0.084 (± 0.265)	0.059 (± 0.204)	0.001 (± 0.009)

overlap between the predicted and ground truth regions, divided by the total number of pixels in both regions. This metric quantifies the similarity between two sets, P and G, where P represents the predicted pixels, and G represents the ground-truth pixels. The Dice score is determined by comparing the overlap between the two sets, denoted as PG (the number of pixels common to both predicted and ground truth sets), with the total number of pixels in P and G.

Tables 3 and 4 present the evaluation of the three segmenters trained on normal glomeruli and tested on normal and segmentally sclerotic glomeruli, respectively. The results show that any of the network models were able to reach state-of-the-art (SOTA) performance, either for testing over normal (considering U-Net, U-Net 3+, and SwinTransformer + U-Net; average mDice of 0.937, 0.931, and 0.935, respectively) or segmentally sclerotic glomerulus (considering U-Net and SwinTransformer + U-Net; average mDice of 0.730 and 0.702, respectively).

As shown in Tables 5 and 6, all three DL models, U-Net, U-Net 3+, and SwinTransformer + U-Net, performed poorly in segmenting globally sclerotic glomeruli, regardless of whether fine-tuning was applied in the WSI segmentation task. Notably, U-Net 3 + and SwinTransformer + U-Net both recorded mDice scores of zero when trained exclusively on normal glomeruli and evaluated on globally sclerotic cases. In contrast, the foundation model SegGPT achieved substantially better results. When provided with six query examples (two per class: normal, segmental, and global), SegGPT reached an average mDice of 0.285, which further improved to 0.428 when using twelve examples (four per class). An illustrative example of the latter setup is shown in Fig. 3. Additionally, Figs. 4 and 5 present cases where SegGPT performed remarkably well and, by contrast, cases where it failed entirely to segment the glomeruli.

Discussion

As shown in Table 2, the number of glomeruli with global sclerosis per WSI is very small. Consequently, model performance on WSIs containing only globally sclerotic glomeruli remains consistently poor for supervised ML models, regardless of whether fine-tuning was applied using HE-stained patches. In most cases, the models fail to even localize these glomeruli, as reflected by Dice scores approaching zero (check Tables 5 and 6). Jiang et al. [9] addressed this limitation by manually cropping individual glomeruli from WSIs and using those patches for training, which simplifies the segmentation task but compromises clinical applicability and generalization to full-slide analysis.

As previously noted, Fig. 3 illustrates an example of the twelve-query setup and provides a qualitative assessment of SegGPT's segmentation performance for globally sclerotic glomeruli at the WSI level, complementing the quantitative results in Table 6. In this evaluation, SegGPT was prompted with six or twelve annotated query examples spanning three glomerular classes (normal, segmental sclerosis, and global sclerosis) and tested on WSIs containing globally sclerotic glomeruli.

Table 6 Mean ± standard deviation of dice scores for globally sclerotic glomerulus segmentation, evaluated on the global sclerosis WSI test set. Columns 3–5 report scores for three networks fine-tuned on the mixed dataset, using pre-trained weights from the normal training set (see section Datasets). Columns 6–7 present results for the SegGPT model using 6 and 12 query examples, which leverage feature representations retrieved from the mixed dataset

Input	Stain	Fine-tuning			SegGPT#6	SegGPT#12
		U-Net	U-Net3+	SwinTransformer + U-Net		
WSI	HE	0.066 (± 0.054)	0.000 (± 0.000)	0.000 (± 0.000)	0.285 (± 0.428)	**0.428 (± 0.258)**
	PAS	0.011 (± 0.009)	0.151 (± 0.039)	0.131 (± 0.041)	0.098 (± 0.288)	**0.366 (± 0.261)**
	ALL	0.041 (± 0.013)	0.124 (± 0.033)	0.034 (± 0.015)	0.198 (± 0.378)	**0.400 (± 0.255)**
Per-crop	HE	0.420 (± 0.142)	0.386 (± 0.108)	0.112 (± 0.110)	0.740 (± 0.329)	**0.742 (± 0.350)**
	PAS	0.448 (± 0.201)	0.461 (± 0.126)	0.224 (± 0.203)	0.566 (± 0.404)	**0.654 (± 0.363)**
	ALL	0.587 (± 0.284)	0.330 (± 0.150)	0.121 (± 0.114)	0.661 (± 0.373)	**0.701 (± 0.354)**

In the HE-stained WSI (Fig. 3a), only one globally sclerotic glomerulus is present. SegGPT successfully detected and segmented it, as shown by the green true-positive mask. The zoomed-in view reveals that the segmentation overlaps well with the ground truth, with only minor areas of false-negative pixels (red) along the glomerular border. This output illustrates the model's capacity to localize isolated sclerotic glomeruli when provided with high-contrast morphological cues and minimal tissue interference. In contrast, the PAS-stained WSI (Fig. 3b) highlights some of the challenges inherent to inference in more complex or ambiguous tissue contexts. The WSI-level output shows two correctly identified glomeruli (green masks), alongside two missed glomeruli (red masks, false negatives) and one spurious detection (yellow mask, false positive). The zoomed patch further illustrates a nuanced case where one glomerulus was entirely missed, while the adjacent glomerulus was partially segmented — with a dominant true-positive region but overlapping false-positive and false-negative areas. This example demonstrates how SegGPT can approximate the glomerular boundary but may struggle with precise delineation in densely packed or morphologically ambiguous regions.

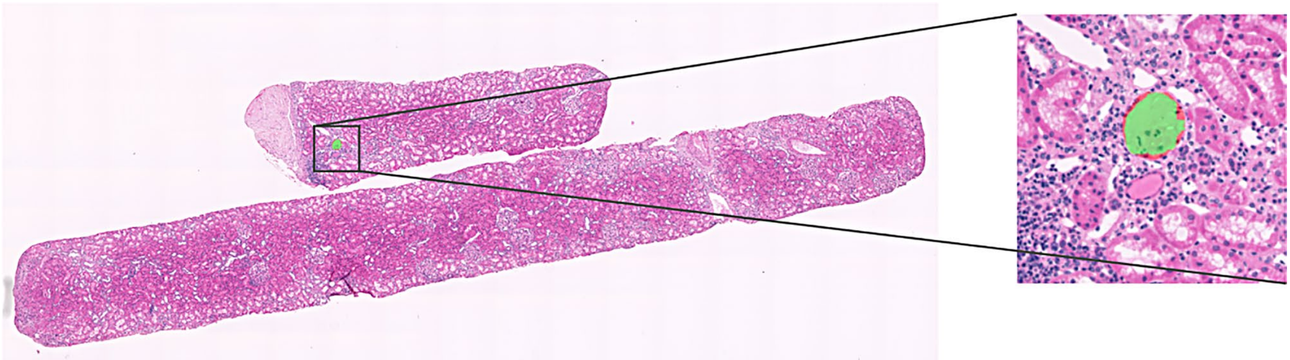
Tables 5 and 6 also present the results of extending our evaluation framework to per-crop inputs specifically targeting globally sclerotic glomeruli. The highest performance was achieved by the SegGPT foundation model when using twelve query examples, with an average Dice score of 0.742. While this result demonstrates strong segmentation capability at the patch level, it does not imply immediate clinical generalization. In the per-crop setting, the glomerulus is already fully contained, centered, and visually prominent, with minimal interference from surrounding renal structures. In real-world diagnostic workflows, models must locate and delineate glomeruli directly within the full-slide context, where variability in staining, background tissue, artifacts, and overlapping structures significantly increases the challenge.

Still in per-crop analyses, U-Net, U-Net 3+, and SwinTransformer + U-Net achieved average Dice scores of 0.587, 0.461, and 0.224, respectively. These models were fine-tuned using weights from the initial framework and trained exclusively on the Mixed dataset, which comprises cropped glomeruli across the three lesion types (see Section Datasets). It is noteworthy that without fine-tuning, all non-foundation models failed to segment globally sclerotic glomeruli, producing Dice scores of zero or near zero.

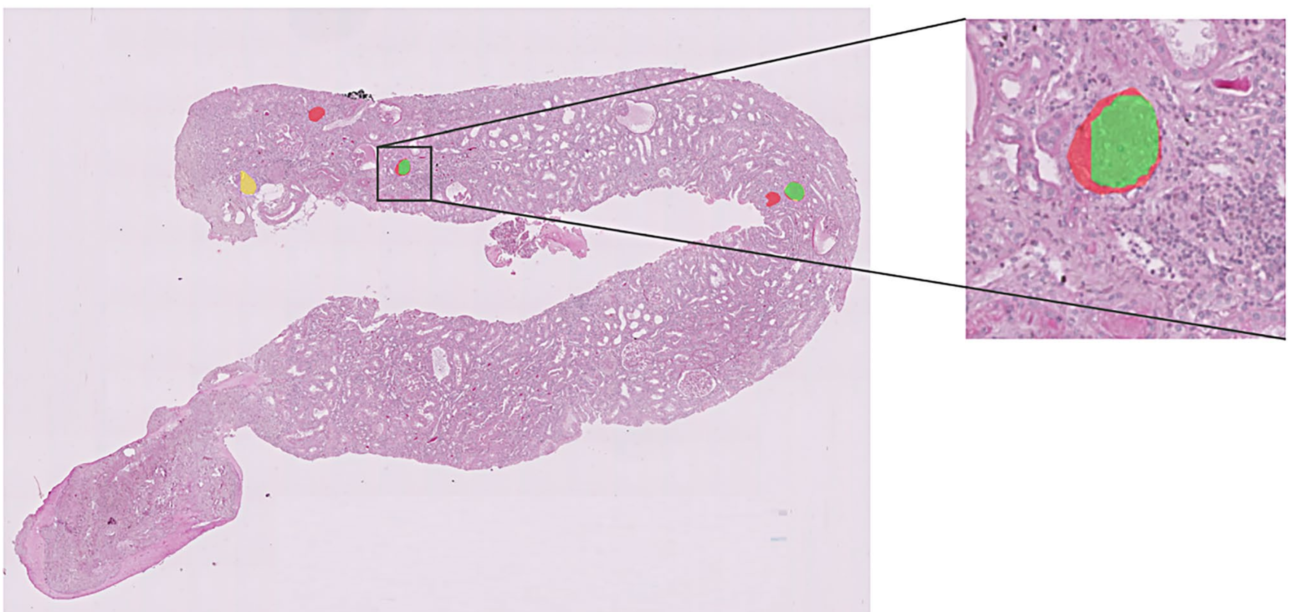
Figures 4 and 5 illustrate the best and worst examples of per-crop segmentation by SegGPT. In Fig. 4, all three staining protocols yielded visually consistent results, with true-positive segmentations clearly highlighted in green, minimal false positives in yellow, and few false negatives in red. In contrast, Fig. 5 presents cases that were marked by a lack of true-positive segmentation, accompanied by false negatives (red) and false positives (yellow), particularly in areas outside the glomerular boundary.

Limitations

A limitation of this study is the intentional underrepresentation of globally sclerotic glomeruli in our dataset. This design choice was made to evaluate how conventional supervised segmentation methods perform when trained on limited examples of this lesion type, thereby mimicking the scarcity of global sclerosis in routine renal pathology. While this allowed us to test the robustness of these models under realistic data constraints, it also means that the reported results may not fully reflect performance in settings with more balanced lesion distributions.



(a)



(b)

Fig. 3 Example outputs from SegGPT's WSI-level segmentation of globally sclerotic glomeruli: (**a**) HE-stained WSI and zoomed patch: The zoomed patch shows detailed true-positive pixels (green) with minimal false negatives (red). (**b**) PAS-stained WSI and zoomed patch: The WSI contains one false positive (yellow mask), two false negatives (red), and two true positives (green masks); the zoomed patch displays one glomerulus with a predominance of true-positive pixels (green), along with false-negative regions (red)

Another limitation concerns the marked performance gap between patch-level and WSI-level segmentation observed with SegGPT. Although the model achieved high accuracy on cropped patches, performance dropped substantially when applied to entire WSIs. This reflects the increased complexity of whole-slide inference, where variability in tissue quality, staining, and the presence of non-target structures can hinder segmentation accuracy.

Finally, the models were not evaluated on any external public dataset. Although this approach is consistent with methodologies adopted in related studies [9, 12, 13], it constrains the assessment of model robustness and generalizability across heterogeneous data sources and scanning conditions.

Conclusions

This study examined glomerular segmentation in both whole-slide images (WSIs) and cropped samples of human kidney tissue, with emphasis on the segmentation of globally sclerotic glomeruli. Four semantic segmentation models were evaluated across three datasets, revealing that while glomeruli with preserved anatomical features can be segmented with high accuracy, the morphological deterioration seen in global sclerosis remains a major obstacle for current methods. Non-foundation architectures were unable to reliably identify these lesions, whereas the foundation model, SegGPT, demonstrated markedly superior performance, achieving a Dice score above 0.37 at the WSI level and approximately 0.74 in the per-crop setting. These results indicate the

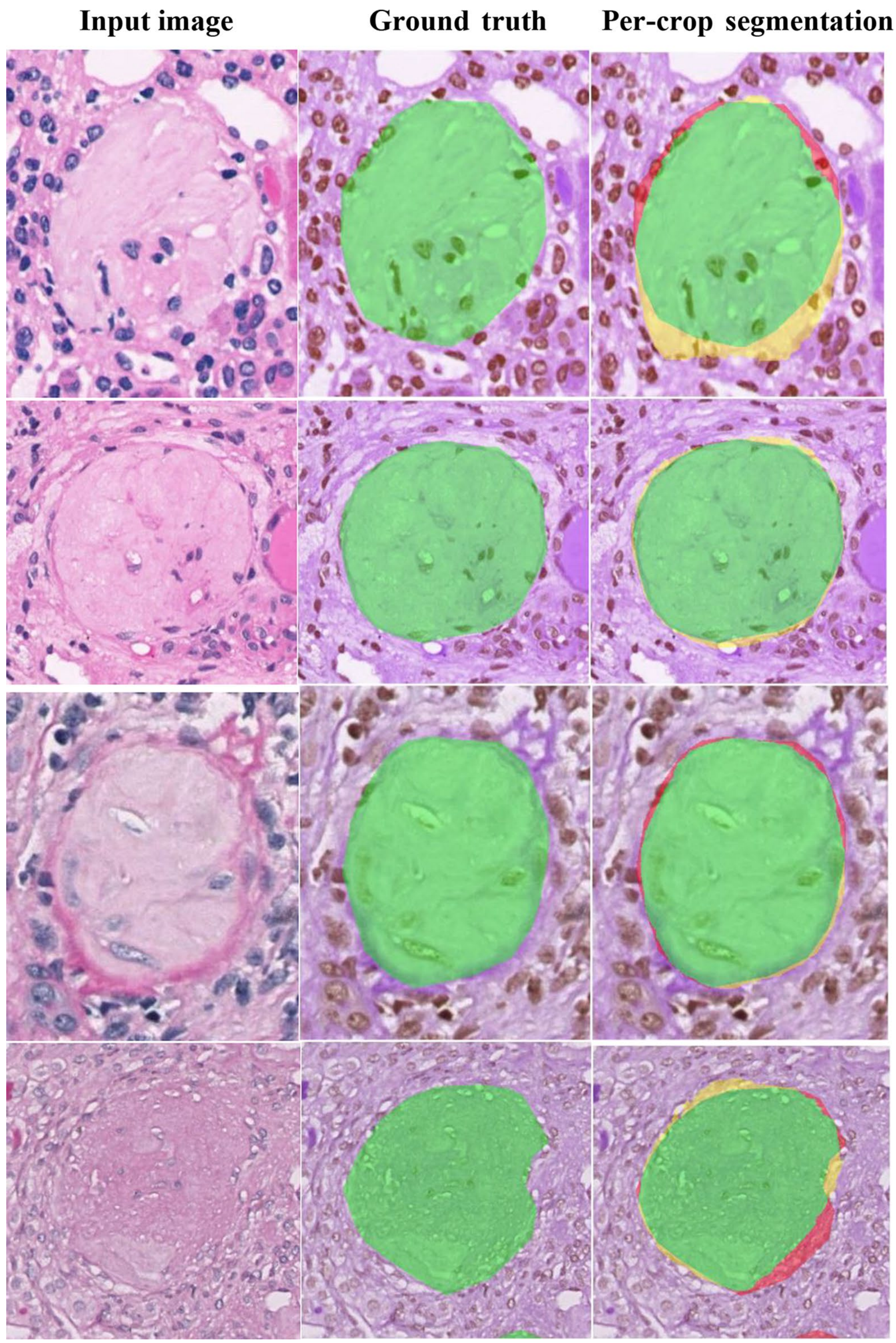


Fig. 4 Examples of best results from SegGPT on the Global Sclerosis test set. From left to right: input image, ground truth annotation, and per-crop segmentation output for global sclerosis. Green masks denote true-positive pixels, yellow masks indicate false positives, and red masks represent false negatives

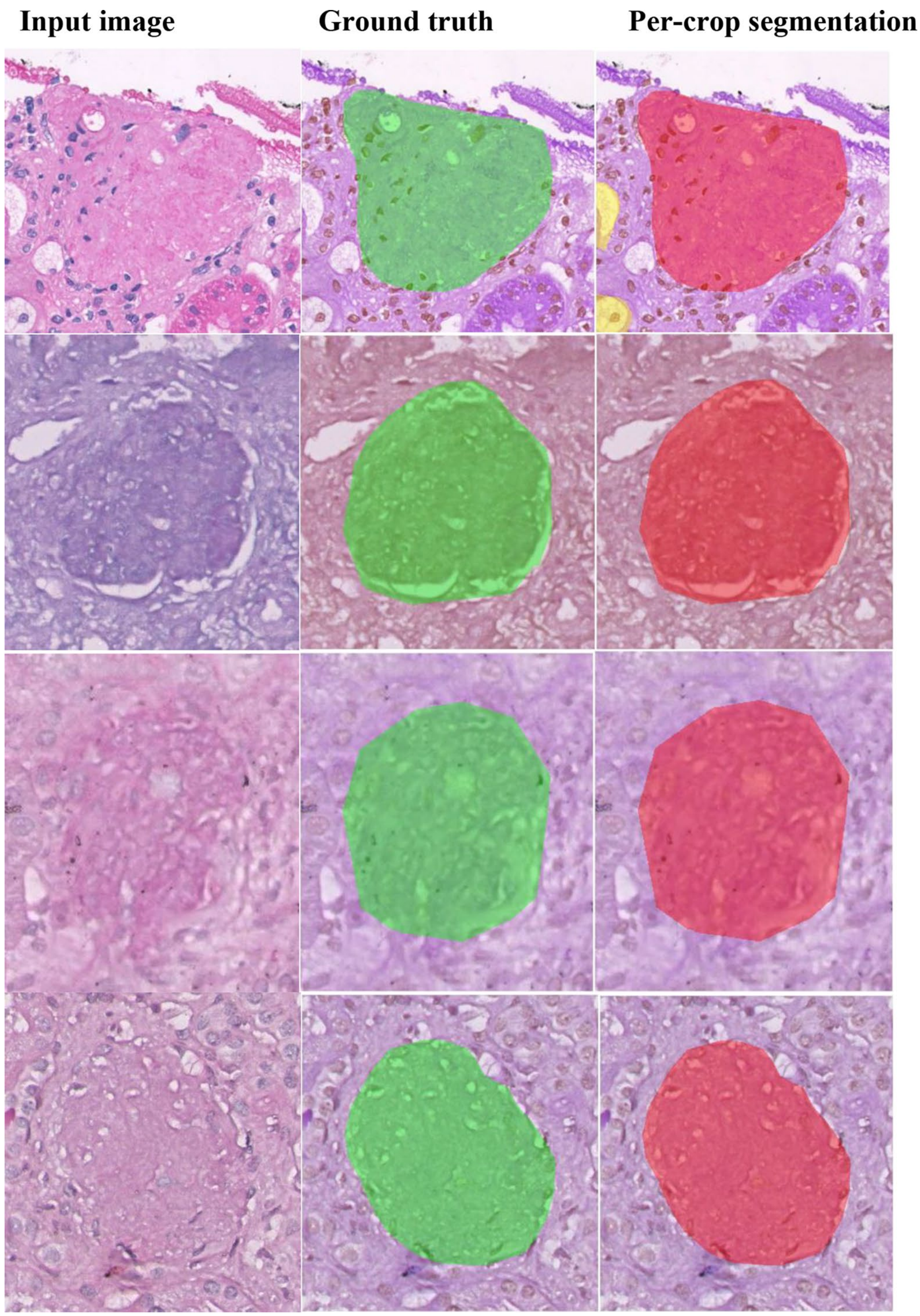


Fig. 5 Examples of missed segmentations by SegGPT on the globally sclerotic test set. From left to right: input image, ground truth annotation, and SegGPT per-crop segmentation output for global sclerosis. Red masks indicate false-negative pixels, and yellow masks denote false positives. The first two rows correspond to HE-stained images, and the last two rows to PAS-stained images

potential of foundation models to address complex renal pathology patterns, but also underscore that current capabilities are still insufficient for direct integration into unsupervised diagnostic workflows.

For clinical applicability, achieving consistent and high-precision segmentation at the WSI level is critical, as diagnostic interpretation depends on the global distribution and context of lesions. Addressing this challenge will require future work to incorporate larger and more diverse training datasets that include full-slide images, adopt progressive fine-tuning strategies from clean patches to complex WSIs, and explore spatial priors or hierarchical inference mechanisms to maintain local accuracy while capturing broader contextual information. Finally, integrating post-processing techniques might further enhance segmentation quality by refining mask boundaries and reducing artifacts.

Abbreviations

SegGPT	Segmentation generative pre-trained transformer
WSIs	Whole slide images
HE	Hematoxylin and eosin
PAS	Periodic acid-Schiff
PAMS	Periodic acid-methenamine silver
DL	Deep learning
ML	Machine learning
mIoU	Mean Intersection over Union
mDice	Mean Dice similarity coefficient
FSGS	Focal segmental glomerulosclerosis
HuBMAP	Human BioMolecular Atlas Program
BCE	Binary cross-entropy
SOTA	State-of-the-Art

Acknowledgements

The authors would like to extend their sincere gratitude to the Oswaldo Cruz Foundation (FIOCRUZ) - Gonçalo Moniz Institute, Bahia, Brazil, for providing access to the renal biopsy images and essential technical support for the development of this work. We also acknowledge the fundamental support of the IVISION Lab, Federal University of Bahia (UFBA), Bahia, Brazil, for the development of the DL model and the technological processing of subsequent stages. Their expertise and infrastructure were indispensable for the execution of this research.

Author contributions

Luiz Souza - design of the work, analysis, interpretation of the data, creation of new software used in the work, have drafted the work, substantively revised. Jefferson Silva - design of the work, analysis, interpretation of the data, creation of new software used in the work, have drafted the work, substantively revised. Marcelo Mendonça - design of the work, analysis, interpretation of the data, creation of new software used in the work, have drafted the work, substantively revised. José Nathan - interpretation of the data, substantively revised. Angelo Duarte - interpretation of the data, substantively revised. Pinaki Sarder - interpretation of the data, substantively revised. Washington LC dos-Santos - interpretation of the data, substantively revised. Luciano Oliveira - design of the work, analysis, interpretation of the data, creation of new software used in the work, have drafted the work, substantively revised.

Funding

Luiz Souza has scholarships from FAPESB, grant No. TO-BOL0660/2018. Angelo Duarte has support from Universidade Estadual de Feira de Santana (UEFS), grant FINAPESQ TO 074/2021 and from Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) grant No. TO PET-0017/2024. Washington LC dos-Santos and Luciano Oliveira are sponsored by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under grants 406141/2023 and 301789/2025-8, respectively.

Data availability

The data generated in this study will not be deposited in a public repository due to privacy and confidentiality concerns related to sensitive patient information from kidney biopsies. Additionally, the dataset comprises proprietary histopathological images governed by strict access agreements that prevent unrestricted distribution. Fiocruz currently restricts public access to patient data to safeguard privacy. Moreover, the complexity of the data may lead to misinterpretation without adequate context. However, to support transparency and collaboration, we are willing to share the data with qualified researchers upon request, contingent on institutional approvals and adherence to ethical guidelines. This approach allows us to balance ethical considerations with research integrity and collaborative advancement.

Declarations

Ethics approval and consent to participate

This work was conducted following resolution No. 466/12 of the Brazilian National Health Council and in accordance with the Declaration of Helsinki. The images (including those shown in this paper) were separated from other patients' data to preserve confidentiality. No data presented herein allows patient identification. Informed consent was waived by the Institutional Review Board for Research Studies with Human Subjects. All the procedures were approved by the Ethics Committee for Research Involving Human Subjects of the Gonçalo Moniz Institute from the Oswaldo Cruz Foundation (CPqGM/FIOCRUZ), Protocols No. 188/09 and No. 1817574.

Consent for publication

All authors agree to publish the article.

Competing interests

We hereby inform that there is a potential conflict of interest related to the work of the authors of the scientific article titled "A Deep Learning-Based Approach for Glomeruli Instance Segmentation from Multistained Renal Biopsy Pathologic Images" (Lei Jiang, Wenkai Chen, Bao Dong, Ke Mei, Chuang Zhu, Jun Liu, Meishun Cai, Yu Yan, Gongwei Wang, Li Zuo, and Hongxia Shi, The American Journal of Pathology, 2021191(8):1431–1441. This conflict arises from our conducting a comparative analysis and raising questions regarding the methodological aspects of the aforementioned study, which could influence the impartial evaluation of our work in case of being reviewed by one of those authors. We emphasize, however, that all analyses and critiques were conducted with the utmost objectivity and transparency, based on technical and scientific evidence.

Author details

[1]IVISION Lab, Universidade Federal da Bahia (UFBA), Salvador, Bahia, Brazil
[2]Universidade Federal do Maranhão (UFMA), Balsas, Maranhão, Brazil
[3]Instituto Federal do Maranhão (IFMA), Grajaú, Maranhão, Brazil
[4]Instituto Federal da Bahia (IFBA), Vitória da Conquista, Bahia, Brazil
[5]Universidade Estadual de Feira de Santana (UEFS), Feira de Santana, Bahia, Brazil
[6]Fundação Oswaldo Cruz (FIOCRUZ), Instituto Gonçalo Moniz, Salvador, Bahia, Brazil
[7]University of Florida (UF), Miami, FL, USA

Received: 17 February 2025 / Accepted: 8 September 2025
Published online: 29 September 2025

References

1. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol. 2019;20(5):e253–61.
2. Della Mea V, Demichelis F, Viel F, Dalla Palma P, Beltrami CA. User attitudes in analyzing digital slides in a quality control testbed: A preliminary study. Comput Methods Programs Biomed. 2006;82(2):177–86.
3. Bayramoglu N, Kannala J, Heikkila J. Deep learning for magnification independent breast cancer histopathology image classification. IEEE International Conference on Pattern Recognition (ICPR). 2016:2440–2445.
4. de Bel T, Hermsen M, Smeets B, Hilbrands L, van der Laak J, Litjens G. Automatic segmentation of histopathological slides of renal tissue using deep learning. Med Imaging: Digit Pathol Int Soc Opt Photonics. 2018;10581:285–90.
5. Marsh JN, Matlock MK, Kudose S, Liu T-C, Stappenbeck TS, Gaut JP, Swamidass SJ. Deep learning global glomerulosclerosis in transplant kidney frozen sections. IEEE Trans Med Imaging. 2018;37(12):2718–28.
6. Hermsen M, de Bel T, Boer MD, Steenbergen EJ, Kers J, Florquin S, Roelofs JJ, Stegall MD, Alexander MP, Smith BH, Hilbrands LB, Laak JAWMV. Deep learning–based histopathologic assessment of kidney tissue. J Am Soc Nephrol. 2019;30(10):1968–79.

7. Bueno G, Fernandez-Carrobles MM, Gonzalez-Lopez L, Deniz O. Glomerulosclerosis identification in whole slide images using semantic segmentation. Comput Methods Programs Biomed. 2020;184:1–10.
8. Altini N, Cascarano GD, Brunetti A, Marino F, Rocchetti MT, Matino S, Venere U, Rossini M, Pesce F, Gesualdo L, Bevilacqua V. Semantic segmentation framework for glomeruli detection and classification in kidney histological sections. Electronics. 2020;9(3):1–15.
9. Jiang L, Chen W, Dong B, Mei K, Zhu C, Liu J, Cai M, Yan Y, Wang G, Zuo L, Shi H. A deep learning-based approach for glomeruli instance segmentation from multistained renal biopsy pathologic images. Am J Pathol. 2021;191(8):1431–41.
10. Davis RC, Li X, Xu Y, Wang Z, Souma N, Sotolongo G, Bell J, Ellis M, Howell D, Shen X, Lafata KJ, Barisoni L. Deep learning segmentation of glomeruli on kidney donor frozen sections. J Med Imaging. 2021;8(6):067501–067501.
11. Gallego J, Swiderska-Chadaj Z, Markiewicz T, Yamashita M, Gabaldon MA, Gertych A. A u-net based framework to quantify glomerulosclerosis in digitized Pas and h&e stained human tissues. Comput Med Imaging Graph. 2021;89:101865.
12. Yu L, Yin M, Deng R, Liu Q, Yao T, Cui C, Guo J, Wang Y, Wang Y, Zhao S, Yang H, Huo Y. Glo-In-One-v2: holistic identification of glomerular cells, tissues, and lesions in human and mouse histopathology. 2024:1–27. https://www.arxiv.org/abs/2411.16961.
13. Wang X, Zhang J, Xu Y, Huang Y, Ming W, Jiao Y, Liu B, Fan X, Xu J. Glo-net: A dual task branch based neural network for multi-class glomeruli segmentation. Comput Biol Med. 2025;186:109670.
14. Yue J, Yao T, Deng R, Lu S, Guo J, Liu Q, Xiong J, Yin M, Yang H, Huo Y. Glo-Finder: AI-empowered QuPath plugin for WSI-level glomerular detection, visualization, and curation. J Pathol Inf. 2025;17:100433.
15. Mateos-Aparicio-Ruiz I, Pedraza A, Becker JU, Altini N, Salido J, Bueno G. GNCnn: A QuPath extension for glomerulosclerosis and glomerulonephritis characterization based on deep learning. Comput Struct Biotechnol J. 2025;27:35–47.
16. Deng R et al. Kpis 2024 challenge: Advancing glomerular segmentation from patch-to slide-level. arXiv preprint arXiv:2502.07288. 2025.
17. Yang F, He Q, Wang Y, Zeng S, Xu Y, Ye J, He Y, Guan T, Wang Z, Li J. Unsupervised stain augmentation enhanced glomerular instance segmentation on pathology images. Int J Comput Assist Radiol Surg. 2025;20(2):225–36.
18. Liu Y, He Q, Duan H, Shi H, Han A, He Y. Using sparse patch annotation for tumor segmentation in histopathological images. Sensors. 2022;22(16):6053.
19. Gadermayr M, Dombrowski A-K, Klinkhammer BM, Boor P, Merhof D. Cnn cascades for segmenting sparse objects in Gigapixel whole slide images. Comput Med Imaging Graph. 2019;71:40–8.
20. Jha A, Yang H, Deng R, Kapp ME, Fogo AB, Huo Y. Instance segmentation for whole slide imaging: end-to-end or detect- then-segment. J Med Imaging. 2021;8(1):1–16.
21. Jayapandian CP, Chen Y, Janowczyk AR, Palmer MB, Cassol CA, Sekulic M, Hodgin JB, Zee J, Hewitt SM, O'Toole J, Toro P, Sedor JR, Barisoni L, Madabhushi A. Development and evaluation of deep learning–based segmentation of histologic structures in the kidney cortex with multiple histologic stains. Kidney Int. 2021;99(1):86–101.
22. Bouteldja N, Klinkhammer BM, Bu¨low RD, Droste P, Otten SW, Stillfried SFV, Moellmann J, Sheehan SM, Korstanje R, Menzel S, Bankhead P, Mietsch M, Drummer C, Lehrke M, Kramann R, Floege J. Deep learning–based segmentation and quantification in experimental kidney histopathology. J Am Soc Nephrol. 2021;32(1):52–68.
23. Silva J, Souza L, Chagas P, Calumby R, Souza B, Pontes I, Duarte A, Pinheiro N. Santos WL-d, Oliveira L. Boundary-aware glomerulus segmentation: toward one-to-many stain generalization. Comput Med Imaging Graph. 2022;100:102104.
24. Souza L, Silva J, Chagas P, Duarte A, Santos WL-d, Oliveira L. Mouse-to-human transfer learning for glomerulus segmentation. Comput Methods Biomech Biomedical Engineering: Imaging Visualization. 2023;11(5):1996–2005.
25. Altini N, Rossini M, Turkevi-Nagy S, Pesce F, Pontrelli P, Prencipe B, Berloco F, Seshan S, Gibier J-B, Dorado AP, Bueno G, Peruzzi L, Rossi M, Eccher A, Li F, Koumpis A, Beyan O, Barratt J, Mohan HQVC, Nguyen HV, Cicalese PA, Ernst A, Gesualdo L, Bevilacqua V, Becker JU. Performance and limitations of a supervised deep learning approach for the histopathological Oxford classification of glomeruli with Iga nephropathy. Comput Methods Programs Biomed. 2023;242:107814.
26. Deng R, Liu Q, Cui C, Yao T, Long J, Asad Z, Womick RM, Zhu Z, Fogo AB, Zhao S, Yang H, Huo Y. Omni-seg: A scale-aware dynamic network for renal pathological image segmentation. IEEE Trans Biomed Eng. 2023;70(9):2636–44.
27. Deng R, Liu Q, Cui C, Yao T, Yue J, Xiong J, Yu L, Wu Y, Yin M, Wang Y, Zhao S, Tang Y, Yang H, Huo Y. Prpseg: Universal proposition learning for panoramic renal pathology segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2024:11736–11746.
28. Deng R, Liu Q, Cui C, Yao T, Xiong J, Bao S, Li H, Yin M, Wang Y, Zhao S, Tang Y, Yang H, Huo Y. Hats: Hierarchical adaptive taxonomy segmentation for panoramic pathology image analysis. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer. 2024:155–166.
29. Trin H, Tran K, Nguyen N, Cao T, Nguyen B, CovHuSeg. An enhanced approach for kidney pathology segmentation. 2024. https://www.arxiv.org/abs/2411.18893.
30. Haas M, Seshan SV, Barisoni L, Amann K, Bajema IM, Becker JU, Joh K, Ljubanovic D, Roberts ISD, Roelofs JJ, Sethi S, Zeng C, Jennette JC. Consensus definitions for glomerular lesions by light and electron microscopy: recommendations from a working group of the renal pathology society. Kidney Int. 2020;98(5):1120–34.
31. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer Assisted Intervention. Springer. 2015:234–241.
32. Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, Han X, Chen Y-W, Wu J. Unet 3+: a full-scale connected Unet for medical image segmentation. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2020:1055–1059.
33. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:10012–10022.
34. Wang X, Zhang X, Cao Y, Wang W, Shen C, Huang T, Seggpt. Towards segmenting everything in context. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2023:1130–1140.
35. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S, Pytorch. An imperative style, high-performance deep learning library. Adv Neural Inf Process Syst. 2019;32.
36. Borner K, deSouza A, Vuppalanchi D, Holland R, Allam S, Drury C. Hubmap - hacking the kidney. 2021. Available at: https://www.kaggle.com/c/hubmap-kidney-segmentation. Accessed June 05, 2023.
37. Yi-de M, Qing L, Zhi-Bai Q. Automated image segmentation using improved PCNN model based on cross-entropy. Proceedings of the IEEE International Symposium on Intelligent Multimedia, Video and Speech Processing. 2004:743–746.
38. Berman M, Triki AR, Blaschko MB. The lovász-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:4413–4421.

Publisher's note