



Universidade Federal da Bahia  
Escola Politécnica / Instituto de Computação

Programa de Pós-Graduação em Mecatrônica

**A COMPREHENSIVE STUDY ON  
GLOMERULUS SEGMENTATION WITH  
LIMITED TRAINING DATA IN  
HIGH-RESOLUTION HISTOPATHOLOGICAL  
IMAGES**

Luiz Otávio de Oliveira Souza Júnior

TESE DE DOUTORADO

Salvador  
06 de dezembro de 2024



LUIZ OTÁVIO DE OLIVEIRA SOUZA JÚNIOR

**A COMPREHENSIVE STUDY ON GLOMERULUS  
SEGMENTATION WITH LIMITED TRAINING DATA IN  
HIGH-RESOLUTION HISTOPATHOLOGICAL IMAGES**

Esta TESE DE DOUTORADO foi apresentada ao Programa de Pós-Graduação em Mecatrônica da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Doutor em Mecatrônica.

Advisor: Prof. Dr. Luciano Rebouças de Oliveira  
Co-advisor: Prof. Dr. Washington Luis Conrado dos Santos  
Co-advisor: Prof. Dr. Jefferson Fontinele da Silva

Salvador  
06 de dezembro de 2024

Ficha catalográfica elaborada pela Biblioteca Bernadete  
Sinay Neves, Escola Politécnica – UFBA.

S729 Souza Júnior, Luiz Otávio de Oliveira.  
A comprehensive study on glomerulus segmentation with  
limited training data in high-resolution histopathological images /  
Luiz Otávio de Oliveira Souza Júnior. – Salvador, 2024.  
95 f.: il. color.

Orientador: Prof. Dr. Luciano Rebouças de Oliveira.  
Coorientador: Prof. Dr. Washington Luis Conrado dos Santos.  
Coorientador: Prof. Dr. Jefferson Fontinele da Silva.

Tese (doutorado) – Programa de Pós-graduação em  
Mecatrônica - Universidade Federal da Bahia - Escola Politécnica,  
2024.

1. Segmentação semântica de imagem. 2. Biópsias renais. 3.  
Visão computacional. 4. Aprendizado profundo. I. Oliveira,  
Luciano Rebouças de. II. Santos, Washington Luis Conrado dos.  
III. Silva, Jefferson Fontinele da. IV. Universidade Federal da  
Bahia. V. Título.

CDD: 621.39

# TERMO DE APROVAÇÃO

**LUIZ OTÁVIO DE OLIVEIRA SOUZA JÚNIOR**

## **A COMPREHENSIVE STUDY ON GLOMERULUS SEGMENTATION WITH LIMITED TRAINING DATA IN HIGH-RESOLUTION HISTOPATHOLOGICAL IMAGES**

Esta TESE DE DOUTORADO foi julgada adequada à obtenção do título de Doutor em Mecatrônica e aprovada em sua forma final pelo Programa de Pós-Graduação em Mecatrônica da Universidade Federal da Bahia.

Salvador, 06 de dezembro de 2024



Documento assinado digitalmente  
**LUCIANO REBOUCAS DE OLIVEIRA**  
Data: 09/12/2024 17:20:55-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof. Dr. Luciano Rebouças de Oliveira - Orientador  
Universidade Federal da Bahia (UFBA)



Documento assinado digitalmente  
**EDUARDO FURTADO DE SIMAS FILHO**  
Data: 09/12/2024 11:37:53-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof. Dr. Eduardo Furtado de Simas Filho  
Universidade Federal da Bahia (UFBA)

GLADSTON JULIANO  
PRATES  
MOREIRA:03590141603

Digitally signed by GLADSTON  
JULIANO PRATES  
MOREIRA:03590141603  
Date: 2024.12.08 18:48:26 -03'00'

---

Prof. Dr. Gladston Juliano Prates Moreira  
Universidade Federal de Ouro Preto (UFOP)



Documento assinado digitalmente  
**NAYZE LUCENA SANGREMAN ALDEMAN**  
Data: 08/12/2024 12:23:02-0300  
Verifique em <https://validar.iti.gov.br>

---

Profa. Dra. Nayze Lucena Sangreman Aldeman  
Universidade Federal do Delta do Parnaíba (UFDPAr)



Documento assinado digitalmente  
**ISADORA DOS SANTOS LIMA**  
Data: 06/12/2024 22:40:12-0300  
Verifique em <https://validar.iti.gov.br>

---

Profa. Dra. Isadora dos Santos Lima  
Centro Universitário Zarns - Salvador (ZARNS)



*In memory of my father, Luiz Otávio de Oliveira Souza.*





## ACKNOWLEDGEMENTS

I would like to thank my mother, Anaita Monteiro for everything, my sister, Luana Monteiro, for her encouragement. I am deeply grateful to my wife, Patrícia Júlia, who was always by my side on this journey, motivating me, and to my daughter, Izabelle Souza, for her affection and love.

I also thank all friends and colleagues at IvisionLab and PathoSpotter for their help, companionship and “*resenha*” during this period. Special thanks to Paulo Chagas, Marcelo Mendonça, and João Paulo for their invaluable support during the final stretch.

I thank Luciano Rebouças for being much more than just an advisor; I thank him for his work, for his understanding in the difficult times, and for making this learning journey a memorable moment in my life. My sincere thanks also go to Dr. Washington and Jefferson Fontinele, who have been far more than just co-advisors. Finally, I would like to thank the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) for their financial support.



*“Educação não transforma o mundo  
Educação muda as pessoas.  
Pessoas mudam o mundo.”*

—PAULO FREIRE



## RESUMO

A crescente disponibilidade de imagens de lâminas inteiras escaneadas (WSIs) tem expandido a patologia digital, permitindo tanto a tomada de decisão médica quanto a análise computacional diretamente a partir de imagens de alta resolução. O diagnóstico de doenças renais por meio de WSIs depende da análise de estruturas específicas do tecido, e a análise automática dessas estruturas requer a segmentação precisa de componentes essenciais, como glomérulos, túbulos, interstício e vasos. Esta tese foca nos glomérulos, que são fundamentais na avaliação de WSIs após biópsias renais. Essas estruturas são impactadas por lesões associadas a diversas doenças. Na análise de WSIs com técnicas de aprendizado de máquina, os glomérulos costumam ser as primeiras regiões segmentadas para orientar tarefas subsequentes. A cápsula de Bowman (CB) é essencial, pois marca a fronteira entre os componentes glomerulares e o tecido intersticial ao redor. Este trabalho propõe dois estudos para abordar a segmentação semântica de glomérulos em imagens histopatológicas de alta resolução dos rins. No primeiro estudo, investigamos a viabilidade de segmentar glomérulos em WSIs humanas utilizando modelos de aprendizado profundo treinados exclusivamente com dados de camundongos. Camundongos e humanos compartilham diversas semelhanças biológicas, incluindo características genéticas, fisiológicas e estruturais, o que faz com que os camundongos sejam amplamente utilizados como modelos para estudar doenças humanas. Embora essa transferência de conhecimento entre espécies seja consolidada na medicina, ela permanece pouco explorada na patologia computacional, onde as WSIs são objetos primários de pesquisa. Para preencher essa lacuna, avaliamos cinco modelos de segmentação semântica: U-Net, U-Net 3+, Res-U-Net, DeepLabV3+ e MA-Net em conjuntos de dados compostos por 18 WSIs de camundongos e 42 WSIs humanas. O U-Net 3+ apresentou o melhor desempenho na avaliação intra-conjunto, alcançando um DICE médio de 0,930 em imagens de camundongos coradas com HE. Nos dados humanos, o U-Net 3+ também obteve excelente desempenho, com DICES de 0,772, 0,824 e 0,791 nas colorações HE, PAS e PAMS, respectivamente. Além disso, o U-Net 3+ mostrou boa generalização ao ser treinado exclusivamente com dados de camundongos e testado em todo o conjunto de dados humanos, alcançando um DICE de 0,798 para imagens HE. No entanto, o desempenho desses modelos diminuiu ao serem aplicados em colorações diferentes, destacando uma limitação na generalização entre colorações. O segundo estudo aborda os desafios da segmentação de glomérulos sem bordas, afetados por esclerose global. Para isso, desenvolvemos um framework automatizado para recorte e recomposição de patches, eliminando a intervenção manual e simplificando o processo de segmentação. Nossos experimentos mostraram que, embora os modelos convencionais possam alcançar resultados de ponta para glomérulos normais e parcialmente escleróticos, o desempenho deteriora significativamente para glomérulos com esclerose global. A precisão da segmentação nesses casos mostrou alta dependência

da coloração utilizada, com resultados geralmente insatisfatórios. Comparando modelos não-fundamentais (U-Net, U-Net 3+ e SwinTransformer + U-Net), com e sem *fine-tuning*, com o modelo SegGPT, observamos que os modelos não-fundamentais, treinados apenas em glomérulos normais nas colorações HE, PAS e PAMS, apresentaram bom desempenho para glomérulos normais ( $mDice > 0,92$ ) e moderado para glomérulos parcialmente escleróticos ( $mDice > 0,72$ ). No entanto, o desempenho caiu drasticamente para glomérulos com esclerose global ( $mDice > 0,02$ ), com melhorias mínimas mesmo após o *fine-tuning*. Por outro lado, o SegGPT obteve melhoria substancial, atingindo um  $mDice$  superior a 0,37 para glomérulos com esclerose global, utilizando apenas poucas amostras de consulta. Esse resultado destaca o potencial dos modelos fundamentais para enfrentar desafios de segmentação em glomérulos severamente lesionados. Em resumo, os estudos apresentados nesta tese representam um avanço significativo na segmentação de glomérulos em WSIs, oferecendo alternativas eficazes e eficientes para tarefas com dados de treinamento limitados. Nossos achados demonstram o potencial do transfer learning entre camundongos e humanos e o uso de modelos fundamentais para melhorar a segmentação de glomérulos com esclerose.

**Palavras-chave:** segmentação; glomérulo; biópsias renais; aprendizado profundo; imagem de lâmina inteira.

## ABSTRACT

The growing availability of scanned whole slide images (WSIs) has expanded digital pathology, enabling medical decision-making and computational analysis directly from high-resolution images. Kidney disease diagnoses using WSIs rely on the analysis of specific tissue structures, and automatic analysis depends on accurately segmenting key components such as glomeruli, tubules, interstitium, and vessels. This thesis focuses on glomeruli, which are essential in assessing WSIs after kidney biopsies. These structures are impacted by lesions related to various diseases. In machine learning-based WSI analysis, glomeruli are often the first regions segmented to guide subsequent tasks. The Bowman’s capsule (BC) is crucial, marking the boundary between glomerular components and surrounding interstitial tissue. This work proposes two studies aimed at addressing the segmentation of glomeruli in high-resolution kidney histopathological images. In the first study, we investigate the feasibility of segmenting glomeruli in human WSIs using deep-learning models trained exclusively on mouse data. Mice and humans share several biological similarities, including genetic, physiological, and structural characteristics, making mice a common model for studying human diseases. While this cross-species knowledge transfer is well-established in medicine, it remains underexplored in computational pathology, where WSIs serve as primary research objects. To address this gap, we evaluated five semantic segmentation models: U-Net, U-Net 3+, Res-U-Net, DeepLabV3+, and MA-Net, using datasets consisting of 18 mouse WSIs and 42 human WSIs. Among these, U-Net 3+ delivered the best performance in intra-dataset evaluation, achieving an average DICE score of 0.930 on HE-stained mouse images. On human data, U-Net 3+ also excelled, attaining DICE scores of 0.772, 0.824, and 0.791 on HE, PAS, and PAMS stains, respectively. Moreover, U-Net 3+ proved promising generalization when trained solely on mouse data and tested across the entire human dataset, achieving a DICE score of 0.798 on HE-stained images. However, while these models performed well on images within the same staining technique, their performance declined when applied across different stains, highlighting a limitation in cross-stain generalization. The second study focuses on the segmentation challenges posed by borderless glomeruli affected by global sclerosis. We developed an automated framework for patch cropping and stitching, eliminating manual intervention to streamline the segmentation process. Our experiments show that while standard segmentation models can achieve state-of-the-art results for normal and partially sclerotic glomeruli, their performance deteriorates significantly for globally sclerotic glomeruli. Notably, segmentation accuracy for these cases was highly dependent on the staining type and generally remained poor across models. We compared non-foundation models (U-Net, U-Net 3+, and SwinTransformer + U-Net) with and without fine-tuning against the SegGPT foundation model. Non-foundation models, trained exclusively on normal glomeruli with HE, PAS, and PAMS stains, achieved high

performance on normal glomeruli (mDice  $> 0.92$ ) and moderate performance on partially sclerotic glomeruli (mDice  $> 0.72$ ). However, their performance dropped sharply to mDice  $> 0.02$  for globally sclerotic glomeruli, with minimal improvements even after fine-tuning. In contrast, SegGPT demonstrated substantial improvement, achieving a significantly higher mDice score ( $> 0.37$ ) for globally sclerotic glomeruli by leveraging only a few query samples. This result highlights the potential of foundation models in addressing segmentation challenges for glomeruli affected by severe lesions. In summary, the studies presented in this thesis represent a significant step forward in the segmentation of glomeruli in WSIs. Our findings offer a comprehensive analysis of glomerulus segmentation with limited training data, demonstrating the potential of mouse-to-human transfer learning, as well as the use of foundation models to improve segmentation accuracy for glomeruli affected by sclerosis.

**Keywords:** segmentation; glomerulus; kidney biopsies; deep learning; whole slide images.



# CONTENTS

<b>Chapter 1—Introduction</b>	1
1.1 Motivation . . . . .	6
1.2 Goals . . . . .	7
1.2.1 General goal . . . . .	7
1.2.2 Specific goals . . . . .	7
1.3 Contributions . . . . .	7
1.4 Chapter map . . . . .	9
<b>Chapter 2—Background</b>	11
2.1 Image semantic segmentation . . . . .	11
2.2 Evaluation metrics . . . . .	12
2.3 Related work . . . . .	14
2.4 Relation with our work . . . . .	18
2.5 Closure . . . . .	20
<b>Chapter 3—Materials and methods</b>	21
3.1 Data sets . . . . .	22
3.1.1 Lutnick’s - Mouse . . . . .	22
3.1.2 Fiocruz - Human . . . . .	22
3.1.3 HuBMAP - Human . . . . .	23
3.2 Data augmentation . . . . .	24
3.3 Semantic segmenters . . . . .	24
3.3.1 Deep-learning based non-foundation models . . . . .	24
3.3.2 Foundation models . . . . .	30
3.4 Ways to improve glomerulus segmentation . . . . .	30
3.5 Evaluation methodology of the study . . . . .	33
3.5.1 Scientific search in literature . . . . .	33
3.5.2 Methods employed for segmentation . . . . .	33
3.5.3 Segmentation strategy . . . . .	34
3.6 Closure . . . . .	34
<b>Chapter 4—Transfer learning using cross-species data for normal glomerulus segmentation</b>	35
4.1 Background . . . . .	35

4.2	Mouse-to-human transfer learning . . . . .	36
4.3	Materials and methods . . . . .	38
4.3.1	Datasets . . . . .	40
4.3.2	Methods . . . . .	40
4.3.3	Implementation details . . . . .	41
4.4	Results and discussion . . . . .	41
4.4.1	Quantitative analysis . . . . .	41
4.4.2	Qualitative analysis . . . . .	44
4.5	Closure . . . . .	48
<b>Chapter 5—Segmenting bordeless glomerulus impaired by glomerulosclerosis</b>		<b>49</b>
5.1	Background . . . . .	49
5.2	Challenges in segmenting glomerulosclerosis . . . . .	53
5.3	Materials and methods . . . . .	54
5.3.1	Datasets . . . . .	54
5.3.2	Methods . . . . .	54
5.3.3	Implementation details . . . . .	55
5.4	Results and discussion . . . . .	56
5.4.1	Qualitative and quantitative analysis of WSI . . . . .	56
5.4.2	Qualitative and quantitative analysis of cropped glomeruli . . . . .	59
5.5	Closure . . . . .	61
<b>Chapter 6—Discussion and conclusions</b>		<b>63</b>
6.1	Overall results in the experiments . . . . .	63
6.1.1	Mouse-to-human transfer learning . . . . .	63
6.1.2	The problem of segmenting global glomerulosclerosis . . . . .	64
6.2	Strengths and limitations . . . . .	64
6.3	Contribution and future work . . . . .	65
<b>References</b>		<b>69</b>

## LIST OF FIGURES

1.1	An example of a kidney tissue image on a glass slide. The full image has a resolution of $60,914 \times 111,422$ pixels. On the right, a magnified section is shown with a resolution of $4,576 \times 17,118$ pixels at 40x magnification.	2
1.2	General process of kidney biopsy. From left to right: A human being has his kidney examined by biopsy, where a tiny section of kidney tissue is extracted by percutaneous needle puncture. Then, the tissues are cut in 3-6 sections and processed using stain techniques in a glass slide. Finally, the sections are observed for scanning optical microscopy, where a magnification is manually defined (2x, 10x, 20x, 40x). . . . .	3
1.3	Schematic and histological view of the nephron and examples of different stained glomerulus images. This figure includes contributions I made as part of the work proposed in (Silva et al. 2022). (A) - Graphic representation of the nephron: Blood flows from the <b>interlobular artery</b> to the <b>afferent arteriole</b> , enter the glomerular capillaries and leaves the glomerulus through the <b>efferent arteriole</b> . The blood is filtrated in the glomerulus and the filtrate flows through the renal tubules where solutes are exchanged; urine is concentrated and discharged through the renal <b>collecting tubule</b> . (B) - <b>HE</b> stain reveals most of the microscopic structures such as cells, and supporting matrix. (C) - <b>PAS</b> stain highlights the supporting membranes such as Bowman’s capsule and mesangium. (D) - <b>PAMS</b> is a silver stain that enhance visualization of extracellular matrix (membrane and part of interstitial components). (E) - <b>Masson’s trichrome</b> stain highlights interstitial extracellular matrix. The main structures highlighted in the stained glomeruli are: <b>Ar</b> = artery, <b>Art</b> = arteriole, <b>BC</b> = Bowman’s capsule, <b>BS</b> = Bowman’s space, <b>EnC</b> = endothelial cell, <b>EpC</b> = parietal epithelial cell, <b>I</b> = interstitium, <b>Mes</b> = mesangium, <b>P</b> = podocyte, and <b>T</b> = tubule. Image taken from (Silva et al. 2022). . . . .	4
1.4	An example of a whole-slide sub-image without (left) and with (right) manually annotated contours in HE-stained kidney histological tissue. The class/color correspondences: glomerulus (orange), tubules (green), interstitial (yellow) and vessels (blue). . . . .	5
1.5	An example of a glomerulus from the WSI. . . . .	6
2.1	Examples of glomeruli from the WSI dataset. (a) An example of a glomerulus; (b) Binary mask of the glomerulus segmentation; (c) Ground-truth mask of the glomerulus segmentation. . . . .	12

2.2	Visual representation of IoU and Dice evaluation metrics. . . . .	13
2.3	Shows an overview of the approach proposed developed in the scope of the work in (Kawazoe et al. 2022). The parallelograms represent the input or output data, while the rectangles represent the processes. Step 1: Detection of glomeruli, and Step 2: Segmentation of glomerular components into four classes: Bowman’s space, glomerular tuft, crescentic, and sclerotic. Image taken from (Kawazoe et al. 2022). . . . .	16
2.4	Shows an overview of the framework developed in the scope of the work in (Saikia et al. 2023). The proposed framework utilizes MLP-U-Net to generalize the encoder-decoder architecture by incorporating Multi-Layer Perceptron encoders. The CNN embedding module, which consists of downsampling CNN layers that are fed input images and masks. The encoder module comes after the CNN embedding module and is linked to the decoder module by Convolutional and ReLU layers. Image taken from (Saikia et al. 2023). . . . .	17
2.5	Shows an overview of the proposed processing pipeline in the scope of the work in (Gadermayr et al. 2019). Two CNN cascades were proposed for glomerulus segmentation and compared with conventional CNN approaches. Image taken from (Gadermayr et al. 2019). . . . .	18
2.6	Provides an overview of the workflow for the sequential segmentation and classification process, which consists of processing kidney WSIs, extracting patches, resizing images, performing semantic segmentation (SegNet), extracting bounding box, performing classification (AlexNet), and calculating performance metrics, as described in (Bueno et al. 2020). This process includes the steps of processing kidney WSIs, extracting patches, resizing images, performing semantic segmentation (using SegNet), extracting bounding boxes, conducting classification (using AlexNet), and calculating performance metrics, as outlined in (Bueno et al. 2020). The workflow is structured in two stages: (i) segmentation into two classes: Non-glomerulus (black) and glomerulus (white), followed by (ii) classification into two classes: Normal glomerulus and sclerosed glomerulus. Image taken from (Bueno et al. 2020). . . . .	19
2.7	Provides an overview of the proposed framework, as outlined in (Gallego et al. 2021). This framework is designed to segment and classify glomeruli in WSIs of human kidney specimens into normal and sclerosed categories. Image taken from (Gallego et al. 2021). . . . .	20

3.1	FCN developed in the scope of the work in (Long et al. 2015), adapted from a VGG classification network (Simonyan e Zisserman 2014). The blocks $C_1, C_2, C_3, C_4, C_5$ represent the convolutional stages just as they are present in the VGG network. The blocks $F_1, F_2, F_3$ represent the fully connected layers that have been replaced by convolutional layer with a kernel of $1 \times 1$ . The block named pixelwise prediction represents a upsampling layer used to return the convoluted image to the size of the output matrix. Image taken from (Long et al. 2015). . . . .	25
3.2	The U-Net encoder-decoder architecture, developed in (Ronneberger et al. 2015), is illustrated with blue boxes representing multi-channel feature maps, with the number of channels indicated above each box and the x-y size noted in the lower left corner. White boxes denote copied feature maps, while arrows indicate different operations. A key feature of this architecture is the skip connections between the encoder (left) and decoder (right) steps. Image taken from (Ronneberger et al. 2015). . . . .	26
3.3	The U-Net 3+ architecture, developed in (Huang et al. 2020). U-Net 3+ integrates multi-scale features through redesigned skip connections and full-scale deep supervision, producing a segmentation map that is both position-aware and boundary-enhanced. Image adapted from (Huang et al. 2020). . . . .	27
3.4	The Res-U-Net architecture, as presented in (Zhang et al. 2018), consists of three primary components: encoding, bridge, and decoding. The encoding component compresses the input image, the bridge links encoding to decoding, and the decoding reconstructs the image for pixel-level segmentation. Each component is built with residual units, featuring two $3 \times 3$ convolutional blocks with batch normalization, ReLU activation, and convolutional layers. Identity mapping connects the input to output within each unit, ensuring efficient information flow and reducing gradient degradation. Image taken from (Zhang et al. 2018). . . . .	28
3.5	The DeepLabV3+ encoder-decoder architecture, developed in (Chen et al. 2018). The encoder step encodes multi-scale contextual information by applying atrous convolution at multiple scales, while the simple yet effective decoder step refines the segmentation results along object boundaries. Image taken from (Chen et al. 2018). . . . .	29
3.6	The architecture of MA-Net network, developed in (Fan et al. 2020). The encoder step includes Res-blocks, consisting of convolution layers and residual connections, to extract high-dimensional feature information, along with the PAB, while the decoder step employs the MFAB to capture channel dependencies, enriching multi-scale semantic information and enhancing network performance. Image taken from (Fan et al. 2020). . . . .	29
3.7	The architecture of SwinTransformer + U-Net developed in (Cao et al. 2022), which is composed of encoder, bottleneck, decoder and skip connections. Encoder, bottleneck and decoder are all constructed based on swin transformer block (Liu et al. 2021). Image taken from (Cao et al. 2022).	31

3.8	Illustration of overall training framework of SegGPT, developed in (Wang et al. 2023). The authors incorporate diverse segmentation data, including part, semantic, instance, panoptic, person, medical image, and aerial image segmentation, and transform them into the same format of images. They adopt a general Painter (Wang et al. 2023) framework with in-context coloring as the training objective and a random coloring scheme for more flexible and generalizable training. Image taken from (Wang et al. 2023).	32
4.1	Examples of two renal WSI and two glomeruli: (a) A mouse WSI; (b) A human WSI; (c) A mouse glomerulus image from Lutnick’s dataset; (d) A human glomerulus image from the Fiocruz dataset. . . . .	37
4.2	Analytical protocol for this study. From (i) to (vi): (i) kidney biopsy samples are collected from human and mouse specimens, whose histological sections are processed, stained and scanned into WSIs; (ii) glomeruli are manually annotated from sections of kidney biopsies; (iii) each section is divided into patches; (iv) five types of architectures were trained with mice glomerular images; (v) the best model of each architecture is selected to predict human glomerulus classes in each patch; (vi) finally, the resulting masks are stitched to compose the final segmented WSI mask. . . . .	39
4.3	Samples of expected results from U-Net, Unet 3+, ResUNet, DeepLabV3+, MA-Net, when trained on Lutnick’s train set, predicting on Fiocruz entire data set. Glomeruli in yellow are false positives (very few examples), in red are false negatives, and in green are true positive. . . . .	45
4.4	Samples of missing detection from U-Net, Unet 3+, ResUNet, DeepLabV3+, MA-Net when trained on Lutnick’s train set, predicting on Fiocruz entire data set. Note that there is small false-positive regions in yellow on PAMS stain predicted with U-Net, DeepLabv3+, and MA-Net. . . . .	46
4.5	Samples of segmentation results by U-Net 3+, when trained on Lutnick’s train set. This nine images present bruised-boundary glomeruli substantially changed by the following lesions: (A) - a cellular crescent filling the Bowman space, (B) - a fibrous crescent and sclerosis of glomerular tufts, and (C) - periglomerular fibrosis. Note that there is two small true-positive glomeruli region on (A) and (B) lesions. . . . .	47
5.1	Examples of glomeruli (green bounding box) from the WSI data sets. (a) A healthy glomerulus (normal), (b) A partially sclerotic glomerulus (segmental sclerosis), and (c) A global sclerotic glomerulus. . . . .	50
5.2	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. . . . .	55

5.3 Some visual results from SegGPT’s WSI segmentation for global sclerosis. (a) HE-stained WSI and a zoomed patch: The WSI segmentation shows the only globally sclerotic glomerulus (green mask) with true-positive pixels. The zoomed patch highlights detailed true-positive pixels (in green) and very few false negatives (in red). (c) PAS-stained WSI and a zoomed patch: The WSI displays a yellow mask representing false-positive pixels, two red masks indicating false-negative pixels, and two green masks highlighting true-positive pixels. The zoomed patch shows two glomeruli: one with missed detection (red mask) and the other containing true-positive (green mask), false-positive, and false-negative pixels, with true-positive pixels predominating. . . . . 58

5.4 Samples of expected results from SegGPT applied on Global Sclerosis’ test set. From left to right: input image, ground truth, and results from SegGPT’s per-crop segmentation for global sclerosis. Yellow masks represent false-positive pixels, red masks indicate false-negative pixels, and green masks show true-positive pixels (best viewed in color). The first two rows are HE-stained images, while the last two rows are PAS-stained images. . . . . 59

5.5 Samples of missing detection from SegGPT predicting on Global Sclerosis’ test set. From left to right: input image, ground truth, and results from SegGPT’s per-crop segmentation for global sclerosis. Yellow masks represent false-positive pixels, red masks indicate false-negative pixels. The first two rows are HE-stained images, while the last two rows are PAS-stained images . . . . . 60





## LIST OF TABLES

2.1	Summary of related work for glomerulus segmentation. Dashes indicate instances where information was unreported. . . . .	15
3.1	Summary of the three datasets, including descriptions of the staining types, number of WSIs, number of glomerulus, and average image dimensions used for training and testing. The staining methods explored include PAS, HE, and PAMS. . . . .	23
4.1	Summary of the data sets. Lutnick’s data set comprises only mouse data, while Fiocruz is composed of human data. . . . .	40
4.2	Intra-dataset results (Dice score), considering five segmentation networks over the train and test data sets described on Table 4.1. . . . .	43
4.3	Comparative results (Dice score) considering training the network models on <i>Lutnick’s mouse train set</i> ( $M_{train}$ ) and on <i>Fiocruz human train set</i> ( $H_{train}$ ), separately. The prediction was performed on <i>Fiocruz human test set</i> ( $H_{test}$ ). The best results in each $M_{train} \rightarrow H_{test}$ line is in bold. . . . .	43
4.4	Comparative results (Dice score) considering training the network models on <i>Lutnick’s mouse train set</i> ( $M_{train}$ ). Prediction was performed on <i>Fiocruz human test set</i> ( $H_{test}$ ) and <i>Fiocruz human train and test sets</i> ( $H_{entire}$ ). . . . .	44
5.1	Summary of related work for glomerulus segmentation. Dashes indicate instances where information was unreported. . . . .	52
5.2	Summary of the characteristics of the datasets. . . . .	54
5.3	Comparison of Dice scores and their standard deviations obtained from three segmentation networks trained on the Normal’s train set. Predictions were performed on both the Normal’s test set and the Sclerosis’ test set as described in Table 5.2 . . . . .	57
5.4	Comparison of Dice scores and their standard deviations obtained from four segmentation networks tested exclusively on global sclerotic glomeruli using both the input WSI and per-crop data from the Sclerosis’ test set. Columns 3-5 list the results for three segmenters without fine-tuning, trained on the Normal’s train set. Columns 6-8 show the results for the same three segmenters with fine-tuning using the pre-trained model from the Normal train’s set and additional training from the Mixed’s dataset. Column 9 lists the results for the SegGPT segmenter, which uses feature representations (without a training set) by querying the Mixed’s dataset. . . . .	57



## ACRONYM LIST

<b>CNN</b>	convolutional neural networks . . . . .	6
<b>FCN</b>	fully convolutional network . . . . .	24
<b>PAS</b>	periodic acid-Schiff . . . . .	2
<b>HE</b>	Hematoxylin and eosin . . . . .	2
<b>PAMS</b>	periodic acid-methenamine silver . . . . .	2
<b>AZAN</b>	Mallory’s trichrome . . . . .	2
<b>TRI</b>	Trichrome . . . . .	3
<b>CKD</b>	chronic kidney disease . . . . .	1
<b>Art</b>	afferent arteriole . . . . .	3
<b>Ar</b>	renal arteries . . . . .	3
<b>EnC</b>	endothelial cells . . . . .	3
<b>VGG</b>	Visual Geometry Group . . . . .	24
<b>Mes</b>	mesangial cells . . . . .	3
<b>BC</b>	Bowman’s capsule . . . . .	3
<b>BS</b>	Bowman’s space . . . . .	3
<b>DL</b>	deep learning . . . . .	6
<b>CNN</b>	convolutional neural network . . . . .	6
<b>SegGPT</b>	Segmentation Generative Pre-trained Transformer . . . . .	7
<b>WSIs</b>	whole slide images . . . . .	5
<b>Dice</b>	Dice similarity coefficient . . . . .	12
<b>IoU</b>	Intersection over Union . . . . .	12
<b>TP</b>	true positives . . . . .	12
<b>TN</b>	true negatives . . . . .	12
<b>FP</b>	false positives . . . . .	12
<b>FN</b>	false negatives . . . . .	12
<b>MLP-U-Net</b>	Multi-Layer Perceptron U-Net . . . . .	15
<b>MLP</b>	Multi-Layer Perceptron . . . . .	15
<b>IgAN</b>	immunoglobulin A nephropathy . . . . .	15

<b>FSGS</b>	focal segmental glomerulosclerosis . . . . .	23
<b>HuBMAP</b>	Human BioMolecular Atlas Program . . . . .	23
<b>NIH</b>	National Institutes of Health . . . . .	23
<b>FFPE</b>	formalin-fixed paraffin-embedded . . . . .	23
<b>CV</b>	computer vision . . . . .	24
<b>CBAM</b>	convolutional block attention module . . . . .	27
<b>PAB</b>	position-wise attention block . . . . .	30
<b>MFAB</b>	multi-scale fusion attention block . . . . .	30
<b>FMs</b>	foundation models . . . . .	30
<b>ML</b>	machine learning . . . . .	30
<b>LBP</b>	local binary pattern . . . . .	37
<b>SVM</b>	support vector machine . . . . .	37
<b>BCE</b>	binary cross-entropy . . . . .	41
<b>MCD</b>	minimal change disease . . . . .	47

## INTRODUCTION

### Contents

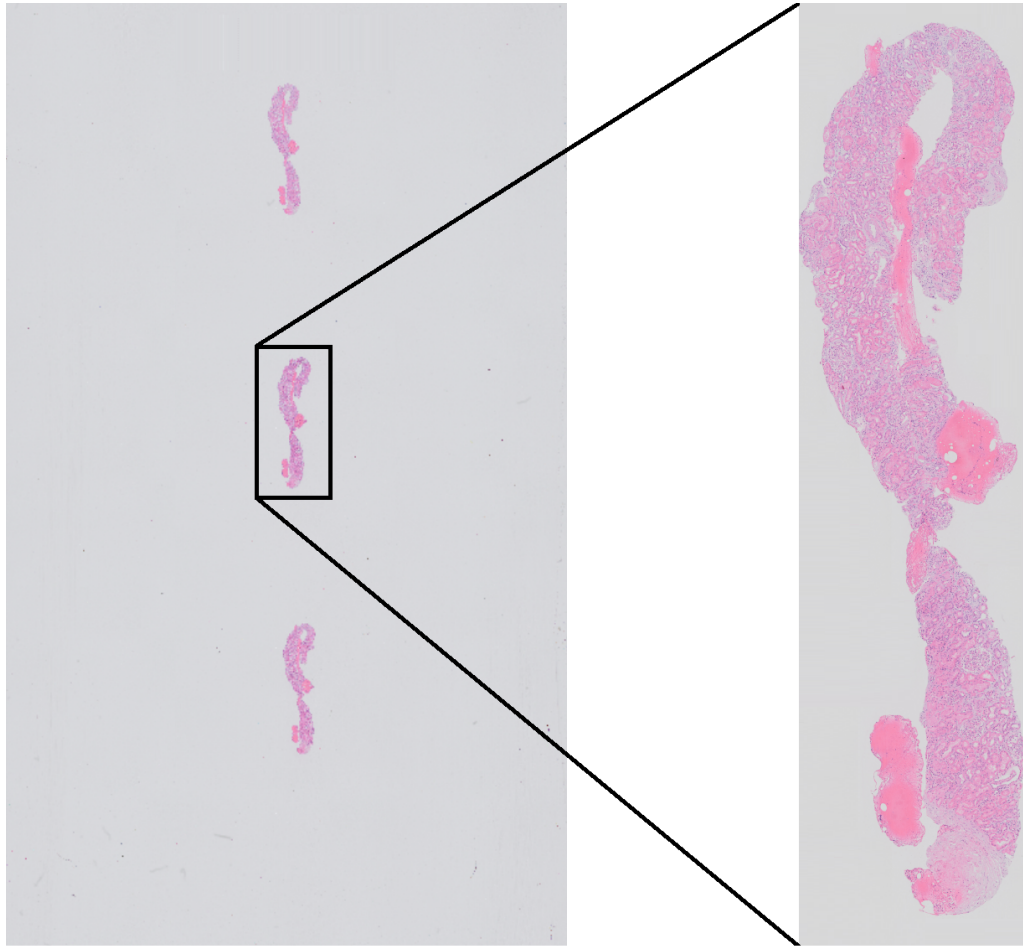
---

<b>1.1</b>	<b>Motivation</b>	<b>6</b>
<b>1.2</b>	<b>Goals</b>	<b>7</b>
1.2.1	General goal	7
1.2.2	Specific goals	7
<b>1.3</b>	<b>Contributions</b>	<b>7</b>
<b>1.4</b>	<b>Chapter map</b>	<b>9</b>

---

Kidneys are vital organs for human body homeostasis that continuously regulate blood pressure and remove waste products from the system (Schwartz et al. 2007). Some damage to the kidney structure can make kidneys unable to remove waste and produce urine. Frequently, kidney disease occurs when the glomeruli, which are formed by tiny blood capillaries, become injured, causing them to lose their filtering capacity. This process of damage in renal function is often undetected in the early stages due to the absence of evident symptoms and the limited evaluation of clinical data. As a result, this silent progression can develop into chronic kidney disease (CKD) if not diagnosed and treated promptly (Levey et al. 2003).

The correct diagnosis of renal disease is crucial for effective treatment to prevent progression to kidney failure and to establish a prognosis. Diagnosis typically relies on clinical and laboratory data, including blood and urine tests. In specific cases, a renal biopsy may be required to confirm the diagnosis (Hogan et al. 2016). Figure 1.1 presents an example of human kidney biopsy sections placed on a glass slide. A kidney biopsy is a medical procedure in which a small fragment of kidney tissue is extracted from the patient's body via percutaneous needle puncture. The sample is then processed for examination using optical, immunofluorescence, and electron microscopy, providing essential information to diagnose, monitor, or guide the treatment of renal failure (Walker et al.

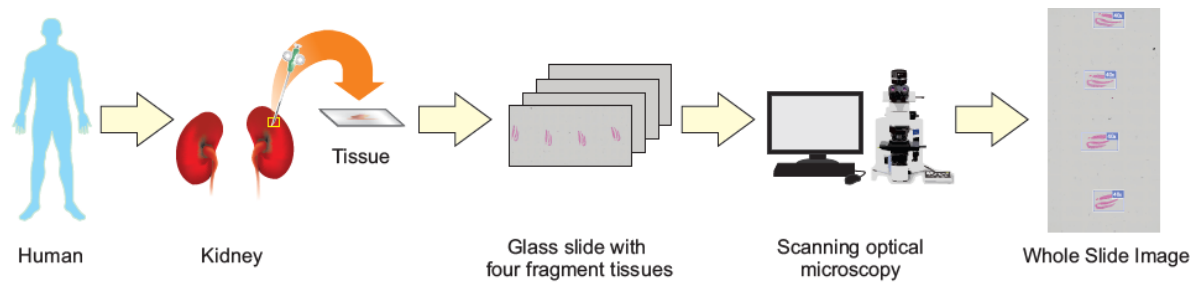


**Figure 1.1** An example of a kidney tissue image on a glass slide. The full image has a resolution of  $60,914 \times 111,422$  pixels. On the right, a magnified section is shown with a resolution of  $4,576 \times 17,118$  pixels at 40x magnification.

2004; Asadzadeh et al. 2019). Figure 1.2 illustrates the medical procedure performed for a kidney biopsy.

Several work groups are engaged in evaluation and standardizing histological staining techniques (Lyon e Horobin 2007; Lyon et al. 1994). Currently, there is a variation of stains in the observed tissue under a microscope by pathologists due to academic medical background or professional experience (Athanzio et al. 2009). The aim is to enhance medical procedures that produce stain substances for analysis of tissue in the microscopic, thereby providing reproducibility in different countries in the field of histopathology (Lyon e Horobin 2007). In pathology laboratories, tissue fragments are commonly stained using one or more of the following techniques: Hematoxylin and eosin (HE), periodic acid-Schiff (PAS), periodic acid-methenamine silver (PAMS), Mallory's trichrome (AZAN), picro-sirius red, Masson's trichrome and Congo red (Alturkistani et al. 2016, Agarwal et al. 2013).

Figure 1.3 (A) depicts the key components of a nephron – the functional unit of the

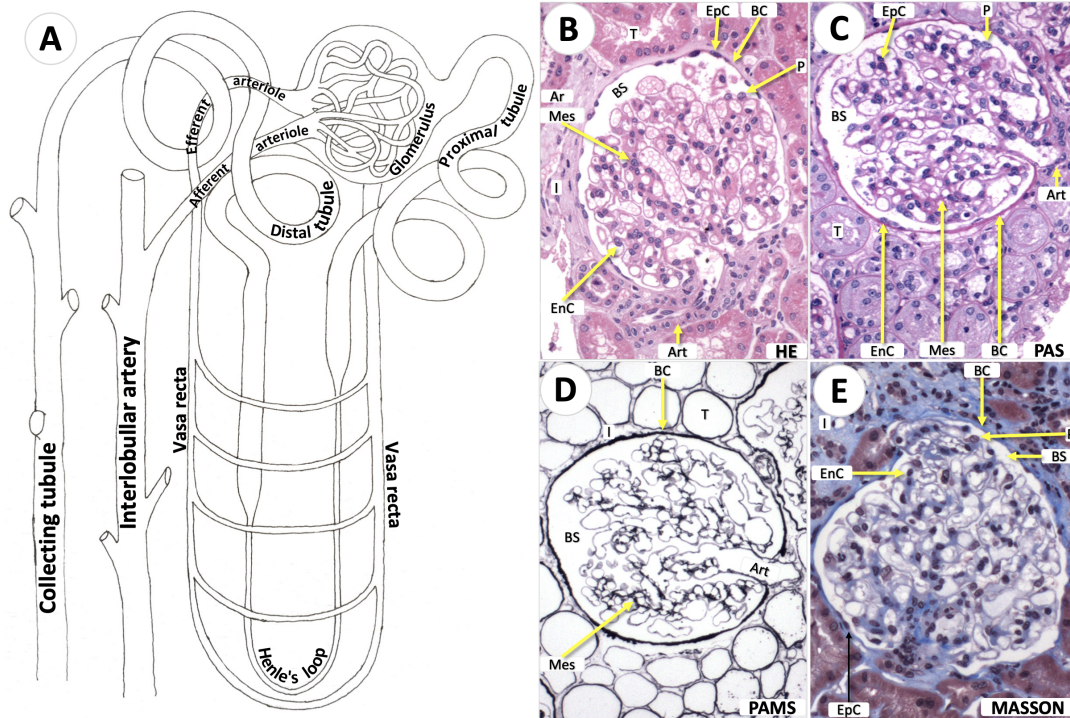


**Figure 1.2** General process of kidney biopsy. From left to right: A human being has his kidney examined by biopsy, where a tiny section of kidney tissue is extracted by percutaneous needle puncture. Then, the tissues are cut in 3-6 sections and processed using stain techniques in a glass slide. Finally, the sections are observed for scanning optical microscopy, where a magnification is manually defined (2x, 10x, 20x, 40x).

kidney – responsible for filtering blood and facilitating the exchange of small molecules, ultimately leading to urine production. Each nephron consists of a glomerulus, along with associated tubules (T), interstitium (I), and blood vessels. Although different diseases can affect one or more nephron components, the glomerulus is the primary target of injury in approximately 90% of renal diseases requiring biopsy (Gesualdo et al. 2004; Polito et al. 2010; Dos-Santos et al. 2017).

The afferent arteriole (Art), a terminal branch of the renal arteries (Ar), gives rise to the glomerulus, a nearly spherical structure composed of intricately interwoven vascular capillary tufts. As blood flows through these capillaries, it is filtered across their walls. The filtrate then passes through the renal tubules, undergoing multiple exchanges of ions and small molecules, resulting in the production of approximately two liters of urine per day, which contains metabolic waste. The glomerular filtration barrier consists of three key components: endothelial cells (EnC), the underlying glomerular basement membrane, and visceral epithelial cells (podocytes, P). These capillary tufts are supported by the mesangium, a central structure composed of mesangial cells (Mes) embedded in a mesangial matrix. The glomerulus is enclosed by Bowman’s capsule (BC), which is made up of fibrous tissue lined internally by a layer of simple squamous epithelium. The space between the parietal and visceral epithelial layers is known as Bowman’s space (BS), where the initial filtrate collects.

Pre-analytical processing of renal biopsies is a fundamental step of a proper morphological analysis. This step includes sample fixation, paraffin-embedding, sectioning in 2-3 micrometers thin slices, and staining by using different techniques. Most nephropathology laboratories rely on four primary staining methods in routine diagnostics: Periodic Acid-Schiff (PAS), Periodic Acid-Methenamine Silver (PAMS), Hematoxylin-Eosin (HE), and Trichrome (TRI) stains. Each technique employs distinct pigments and optimizes their binding to specific tissue structures based on the chemical and physical properties of the components involved. For instance, during HE staining (see Fig. 1.3 (B)) hematoxylin binds to anionic components of the cell nucleus while eosin binds to cationic components of the cell cytoplasm. Therefore, the use of different staining techniques allows a proper visualization of different tissue structures. In the study of renal biopsies,



**Figure 1.3** Schematic and histological view of the nephron and examples of different stained glomerulus images. This figure includes contributions I made as part of the work proposed in (Silva et al. 2022). (A) - Graphic representation of the nephron: Blood flows from the **interlobular artery** to the **afferent arteriole**, enter the glomerular capillaries and leaves the glomerulus through the **efferent arteriole**. The blood is filtrated in the glomerulus and the filtrate flows through the renal tubules where solutes are exchanged; urine is concentrated and discharged through the renal **collecting tubule**. (B) - **HE** stain reveals most of the microscopic structures such as cells, and supporting matrix. (C) - **PAS** stain highlights the supporting membranes such as Bowman's capsule and mesangium. (D) - **PAMS** is a silver stain that enhance visualization of extracellular matrix (membrane and part of interstitial components). (E) - **Masson's trichrome** stain highlights interstitial extracellular matrix. The main structures highlighted in the stained glomeruli are: **Ar** = artery, **Art** = arteriole, **BC** = Bowman's capsule, **BS** = Bowman's space, **EnC** = endothelial cell, **EpC** = parietal epithelial cell, **I** = interstitium, **Mes** = mesangium, **P** = podocyte, and **T** = tubule. Image taken from (Silva et al. 2022).

HE is used to obtain a general view of the tissue. Cell nucleus and cytoplasm are clearly stained contrasting with the light staining of the extracellular matrix. PAS (see Fig. 1.3 (C)) highlights basement membranes and sugar aggregates. There are a variety of silver stain techniques such as Jones methenamine silver and PAMS (see Fig. 1.3 (D)) that clearly delineates details of normal or altered basement membranes. TRI, such as Masson (see Fig. 1.3 (E)) or Azan trichromes, stains collagen and other extracellular matrix components allowing visualization of cell-extracellular matrix relationship (Cathro et al. 2018, Chang et al. 2012).

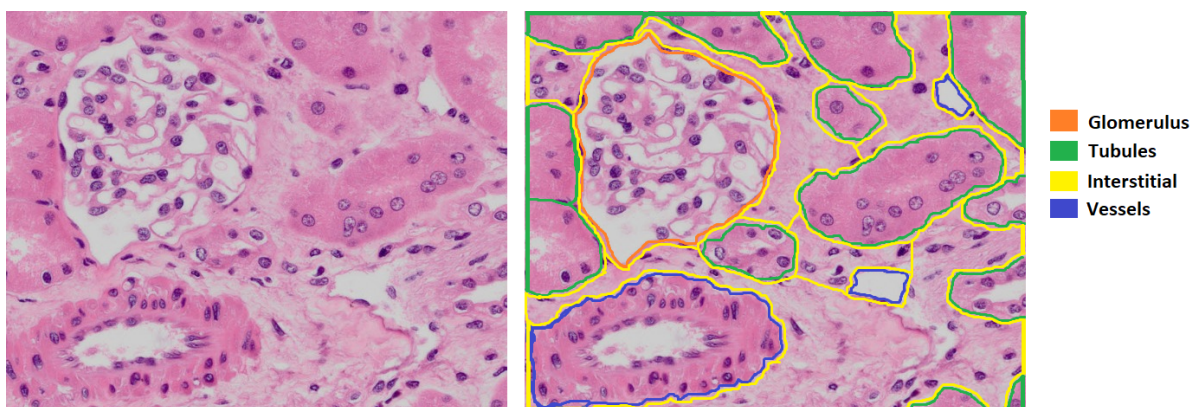
Pathologists commonly analyze four histological tissues by means of a microscope:



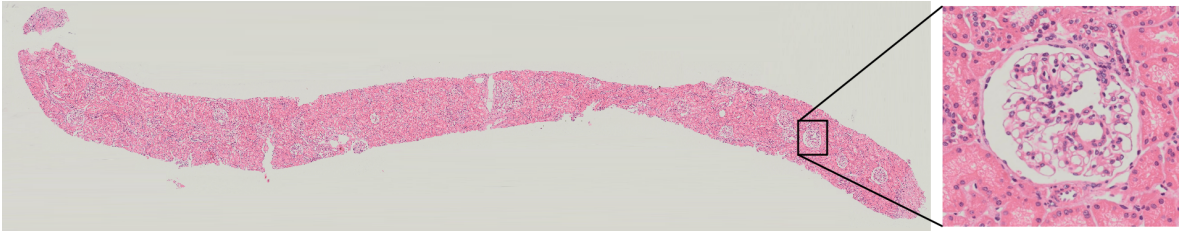
the Glomeruli, tubules, interstitium, and vessels. The **glomeruli** are histological tissues located at the beginning of the nephrons in the kidney, formed by a network of capillaries. The main function of the glomeruli is the filtration of the blood, which leads to urine production. The **tubular** system is comprised of the proximal and distal tubules. The **proximal tubule** consists of convoluted and greater portions, with cells presenting abundant cytoplasm and containing structures for active fluid transport. The **distal tubule** is narrower and shorter than the proximal tubule, and there are more cells per unit area. The **interstitium** tissue contains specialized interstitial cells and connective tissue elements, which provide support to fill spaces around the parenchyma. The renal blood **vessels** are structurally similar to those ones in other parts of the body. Figure 1.4 is an example of tissue renal four structures.

Among different structures in kidney histological tissue, this Ph.D thesis focuses on the glomeruli, a roughly spherical network formed by tiny blood capillaries responsible for blood filtration. Given its primary function, the location and segmentation of glomeruli are valuable information extracted from a kidney whole slide images (WSIs). Figure 1.5 shows an example of a zoomed-in patch of a glomerulus from a WSI. Since localizing glomeruli is time-consuming and error-prone, a promising alternative arises with the development of an automatic glomerular segmentation approach, providing a fast and reliable supportive tool for the pathologists' decision-making pipeline. In this context, we highlight the challenge of gathering a large amount of annotated data, which can be even harder to obtain if one considers human biopsies.

Typically, the segmentation of glomeruli is one of the early tasks performed by machine-learning-based expert systems in computational nephropathology. The primary objective is to provide specialists with a decision-making tool for pre-screening WSIs (*e.g.*, glomerulus counting), thereby streamlining subsequent automated tasks (*e.g.*, glomerular lesion classification). Many studies address the problem of glomerulus segmentation (Bel et al. 2018; Marsh et al. 2018; Gadermayr et al. 2019; Hermsen et al. 2019; Bueno et al. 2020; Altini et al. 2020;



**Figure 1.4** An example of a whole-slide sub-image without (left) and with (right) manually annotated contours in HE-stained kidney histological tissue. The class/color correspondences: glomerulus (orange), tubules (green), interstitial (yellow) and vessels (blue).



**Figure 1.5** An example of a glomerulus from the WSI.

Jha et al. 2021; Jiang et al. 2021; Jayapandian et al. 2021; Bouteldja et al. 2021; Davis et al. 2021; Gallego et al. 2021; Altini et al. 2023). Most of these studies do not differentiate between types of glomeruli, and the performance of the deep learning (DL) models used for segmentation is evaluated by disregarding the various lesions that can affect the glomeruli. Consequently, the performance of the proposed DL models may be overestimated, especially when considering their translation to routine clinical practice. Nonetheless, DL pipelines still require large amounts of data to train the model, either performing it from scratch or fine-tuning it to a target domain.

These studies explore the use of the deep learning models (non-foundation models), and in recent years, the emergence of foundation models has gained significant attention (Chen e Sun 2023; Chen et al. 2024). **Non-foundation models** are deep learning models that are developed without relying on a foundational or “base” model. They are typically trained from scratch or with specific data, without leveraging pre-existing knowledge from a broad task dataset. These models are often designed to address a particular problem and may require substantial training to adapt to different tasks (Chen e Sun 2023). **Foundation models**, on the other hand, are large-scale machine learning models trained on vast datasets, allowing them to perform a variety of tasks. They are often capable of handling different tasks without the need for complete retraining. These models can be fine-tuned for specific applications using techniques like transfer learning, offering versatility across multiple domains, such as natural language processing and computer vision (Chen et al. 2024).

## 1.1 MOTIVATION

Segmenting glomeruli in gigapixel histopathological images is one of the main challenges in computer vision applied in the medical field (Marsh et al. 2018; Gadermayr et al. 2019). Segmentation methods leverage various characteristics to describe the glomeruli to be segmented. Generally, these characteristics do not explicitly account for borderless glomeruli, even though edge delineation is a crucial visual feature for identifying glomeruli in gigapixel histopathological images, especially those affected by global sclerosis lesions.

In this work, to tackle the problem of having limited training data, we explore the use of transfer learning between species (rats and humans) and the problem of segmenting global glomerulosclerosis, aiming to address three main questions: (i) *Is it possible to design an end-to-end convolutional neural network (CNN) for segmentation of glomeruli in high-resolution images?* (ii) *Is it possible to segment human glomerulus from WSIs*

*by using a model trained with only mouse glomerular images? (iii) How do we segment borderless glomeruli?*

To address those research questions, we investigated different segmentation approaches, including deep learning methods and foundation model techniques. Based on our findings, we propose two key investigations: (i) Considering the similarities between mice and humans, the first contribution of our work is the investigation of cross-species compatibility for glomerulus segmentation. We explore whether a DL model trained on mouse data can effectively segment human glomeruli; (ii) the second contribution focuses on the challenge of accurately delineating the boundaries of borderless glomeruli in histopathological images. We assess the robustness of normal glomerulus segmentation and further analyze the critical aspects involved in segmenting borderless glomeruli affected by global sclerosis lesions. This evaluation includes using both traditional deep learning models (non-foundation models) and foundation models (Chen e Sun 2023; Chen et al. 2024).

## 1.2 GOALS

### 1.2.1 General goal

The main goal of this thesis is to conduct a comprehensive study on the segmentation of glomeruli, using limited training data. For that, we aim to investigate glomeruli with clearly defined borders, those with partially defined edges, and glomeruli without distinct borders, in gigapixel histopathological images, using transfer learning, fine-tuning, and few-shot learning techniques.

### 1.2.2 Specific goals

More specifically, other goals are:

- Explore the feasibility of segmenting glomeruli in human WSIs by leveraging transfer learning to train deep learning models exclusively on mouse data;
- Investigate the automated segmentation of globally and partially sclerotic glomeruli using non-foundation (deep learning) models and foundation models;
- Construct annotated datasets provided by Fiocruz to support the research community.

## 1.3 CONTRIBUTIONS

The main contribution of this work is a comprehensive study of glomerulus segmentation in high-resolution histopathological images, focusing on supervised and semi-supervised segmentation tasks. Our approach leverages transfer learning, fine-tuning, and few-shot learning by training on one subset of glomerulus classes and testing on a different subset. We explored a foundation model, called Segmentation Generative Pre-trained Transformer (SegGPT), which employs few-shot learning with a minimal number of images

and achieves remarkable results in segmenting normal glomeruli and those with segmental and global sclerosis. Further details on SegGPT are provided in Chapter 3 and in the article submitted to *Kidney International*, which is currently under review (decision scheduled for January).

1. **SOUZA, L.**; SILVA, J.; MENDONÇA, M.; NATHAN, J.; DUARTE, A.; SARDER, P.; LC-DOS-SANTOS, W.; OLIVEIRA, L. The problem of segmenting global glomerulosclerosis in gigapixel histopathological images: The borderless glomeruli. *Kidney International*, 2024.

Our proposed approach for the supervised segmentation task utilizes training data from glomeruli in one class to predict glomeruli in another class, leveraging efficient methods from the literature to improve segmentation results. The initial contribution of this study lies in the application of deep learning techniques to tackle the challenge of identifying and segmenting glomeruli in gigapixel histopathological images. These efforts have resulted in publications in various conferences and journals.

1. **CHAGAS, P.**; **SOUZA, L.**; ARAÚJO, I.; ALDEMAN, N.; DUARTE, A.; ANGELO, M.; DOS-SANTOS, WL; OLIVEIRA, L. Classification of glomerular hypercellularity using convolutional features and support vector machine. *Artificial Intelligence in Medicine*. 2020 Mar 1;103:101808. [**The first two authors contributed equally to this work.**]
2. **SOUZA, L.**; SILVA, J., CHAGAS, P.; DUARTE, A.; LC-DOS-SANTOS, W.; and OLIVEIRA, L. How feasible is it to segment human glomerulus with a model trained on mouse histology images?. In: *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, 2022.
3. **SOUZA, L.**; SILVA, J.; CHAGAS, P.; DUARTE, A.; SANTOS, W. L.; OLIVEIRA, L.; Mouse-to-human transfer learning for glomerulus segmentation. In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2023.

During the PhD, the following collaborations resulted in publications to conferences and journals:

1. CHAGAS, P.; **SOUZA, L.**; CALUMBY, R.; DUARTE, A; ANGELO, M.; SANTOS, W.; OLIVEIRA, L. Deep-learning-based membranous nephropathy classification and Monte-Carlo dropout uncertainty estimation. In: *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, 2021.
2. CHAGAS, P., G.; **SOUZA, L.**; CALUMBY, R.; PONTES I.; ARAÚJO S.; DUARTE A.; PINHEIRO N.; SANTOS W.; OLIVEIRA L. Toward unbounded open-set recognition to say “I don’t know” for glomerular multi-lesion classification. In: *International Symposium on Medical Information Processing and Analysis (SIPAIM)*, Campinas, 2021.

3. CHAGAS, P.; **SOUZA, L.**, PONTES, I.; CALUMBY, R.; ANGELO, M.; DUARTE, A.; LC-DOS SANTOS, W. and OLIVEIRA, L. Uncertainty-aware membranous nephropathy classification: A Monte-Carlo dropout approach to detect how certain is the model. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2022.
4. SILVA, J.; **SOUZA, L.**, CHAGAS, P.; CALUMBY, R.; SOUZA, B.; PONTES, I.; DUARTE, A.; LC-DOS-SANTOS, W.; and OLIVEIRA, L. Boundary-aware glomerulus segmentation: Toward one-to-many stain generalization. In: *Computerized Medical Imaging and Graphics*, 2022.

## 1.4 CHAPTER MAP

The remainder of this thesis is divided as follows:

- **Chapter 2** provides a concise introduction to segmentation, discussing the fundamental types, key methods, and evaluation metrics.
- **Chapter 3** presents an overview of the methods explored in this comprehensive study, along with a detailed description of the public glomeruli datasets and the private dataset developed at Fiocruz.
- **Chapter 4** outlines our first contribution, focusing on the cross-species compatibility of human and mouse data for glomerulus segmentation.
- **Chapter 5** delves into our second contribution, examining the robustness of segmenting normal glomeruli and the challenges of segmenting borderless glomeruli.
- **Chapter 6** summarizes the key findings of the thesis. Here, we present our final considerations on the strengths and limitations of the proposed approach and suggest perspectives for future research.



**BACKGROUND****Contents**


---

<b>2.1</b>	<b>Image semantic segmentation . . . . .</b>	<b>11</b>
<b>2.2</b>	<b>Evaluation metrics . . . . .</b>	<b>12</b>
<b>2.3</b>	<b>Related work . . . . .</b>	<b>14</b>
<b>2.4</b>	<b>Relation with our work . . . . .</b>	<b>18</b>
<b>2.5</b>	<b>Closure . . . . .</b>	<b>20</b>

---

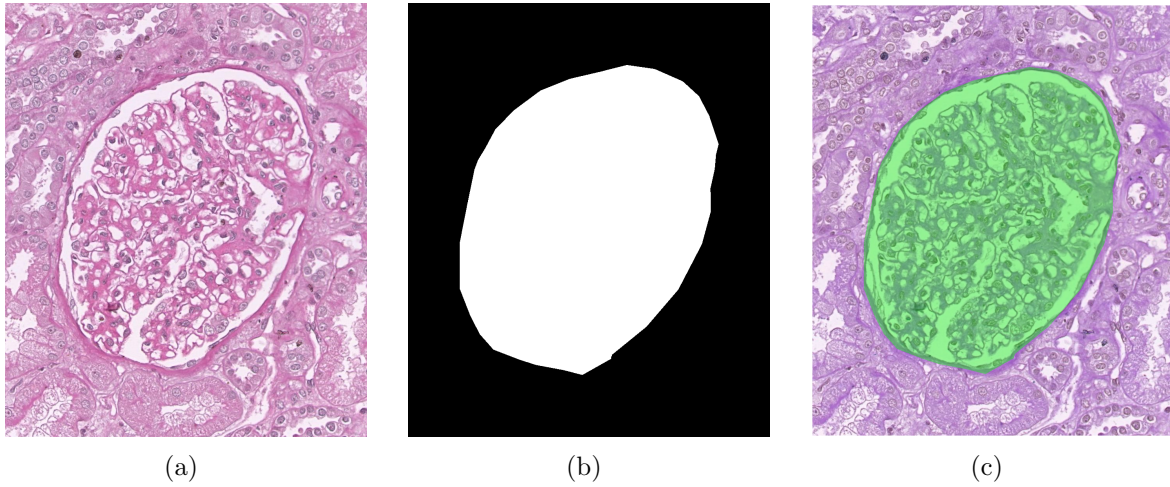
In this chapter, we present the theoretical background that underpins our research. We begin by providing a formal definition of image semantic segmentation. Following this, we explore the metrics commonly used to evaluate its performance. Finally, we discuss related works and their connection to our study.

**2.1 IMAGE SEMANTIC SEGMENTATION**

Image semantic segmentation is the task of classifying each pixel in an image into known structures, which is a semantic label (in our case: glomeruli). Segmenting semantic labels requires pinpointing the contour of structures and thus imposing more complex localization accuracy requirements than other computer vision tasks such as image classification or detection.

**Definition 2.1.1** (Image semantic segmentation). Let  $\mathbb{P}$  be the set of all pixels of the image  $I$ , and  $\mathbb{L}$  a set of labels, where  $\mathbb{L}$  is finite. Image semantic segmentation involves assigning a label  $l \in \mathbb{L}$  to each pixel  $p \in \mathbb{P}$ .

Defining the number of pixels and the semantic classes into which an image should be divided is not a deterministic task; it largely depends on individual perception. In semantic image segmentation, every pixel receives a label. Figure 2.1 illustrates examples of images with and without semantic segmentation of glomeruli. In the segmentation 2.1



**Figure 2.1** Examples of glomeruli from the WSI dataset. (a) An example of a glomerulus; (b) Binary mask of the glomerulus segmentation; (c) Ground-truth mask of the glomerulus segmentation.

(a), an example of a glomerulus, 2.1 (b), the white label represents glomeruli, while the black label denotes other structures, and 2.1 (c), the green label shows ground-truth mask of the glomerulus.

## 2.2 EVALUATION METRICS

The performance of a segmentation model is evaluated using various metrics, with the primary ones being the Dice similarity coefficient (Dice) and Intersection over Union (IoU). These metrics rely on information from a standard confusion matrix, which includes the following components: true positives (TP), where a pixel in the ground truth is positive, and the method correctly identifies it as positive; true negatives (TN), where a pixel in the ground truth is negative, and the method correctly identifies it as negative; false positives (FP), where a pixel in the ground truth is negative, but the method incorrectly identifies it as positive; and false negatives (FN), where a pixel in the ground truth is positive, but the method incorrectly identifies it as negative. These components are used to calculate the difference between the segmentation provided by the ground truth and the prediction masks.

Mean-IOU considers the overlap between the predicted and ground truth masks, divided by the union of the two masks. The Dice score measures twice the overlap between the predicted and ground truth masks, divided by the total number of pixels in both masks. Figure 2.2 depicts both geometric representations of the operations performed by these two metrics. Using the information from the confusion matrix, we can define both metrics, Dice score and IoU, respectively as

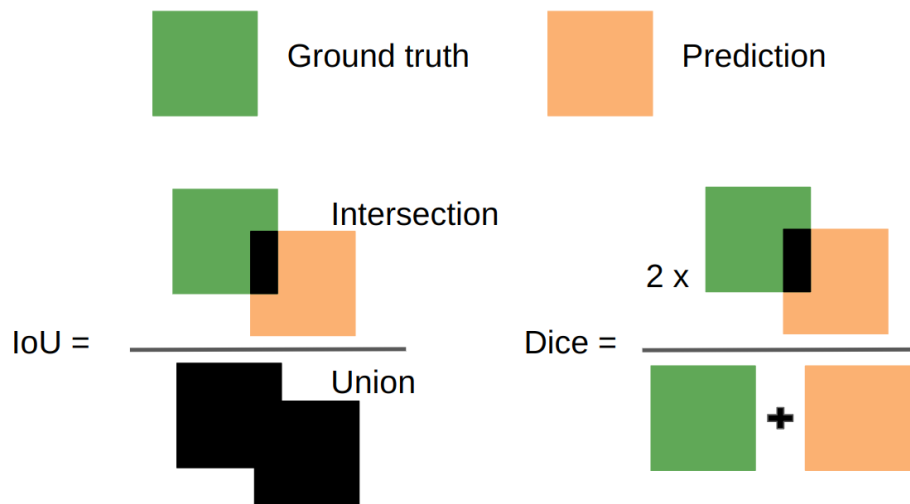
$$Dice(G, P) = \frac{2 \cdot |G \cap P|}{|G| + |P|} = \frac{2TP}{2TP + FP + FN} \quad (1)$$



$$IoU(G, P) = \frac{G \cap P}{G \cup P} = \frac{TP}{TP + FP + FN} \quad (2)$$

Equation (1) shows the Dice metric between two sets  $G$  and  $P$ , where  $G$  is predicted pixels, and  $P$  is the the ground truth. The notation  $|G|$  indicates the number of pixels in the predicted pixels, while  $|P|$  indicates the number of pixels in the ground truth pixels. The term  $|G \cap P|$  denotes the number of pixels common to both the predicted and ground truth pixels.

Equation (2) shows the IoU metric between two sets  $G$  and  $P$ , where  $G$  is predicted pixels, and  $P$  is the the ground truth. The notation  $G$  indicates the number of pixels in the predicted pixels, while  $P$  indicates the number of pixels in the ground truth pixels. The term  $G \cap P$  ndicates the number of pixels common to both the predicted and ground truth sets, while  $G \cup P$  represents the number of pixels in the union of the predicted and ground truth sets.



**Figure 2.2** Visual representation of IoU and Dice evaluation metrics.

The IoU and the Dice are widely used metrics for evaluating segmentation models, each with its own characteristics. IoU, also known as the Jaccard Index, measures the ratio between the intersection and the union of the predicted and ground truth areas. It is more stringent and commonly used in benchmarks but is sensitive to small discrepancies and may not perform well in scenarios with imbalanced classes. On the other hand, the Dice Coefficient calculates the similarity between predictions and ground truth based on the intersection weighted by the size of the areas. It is more tolerant of minor differences and ideal for imbalanced datasets, such as in medical segmentation. However, its lower penalization for significant errors and less frequent use in general benchmarks can be limitations. The choice between them depends on the context and the characteristics of the evaluated data. Dice is typically more sensitive to small classes due to its emphasis

on true positives and false negatives, whereas IoU can be more affected by false positives. Our study employed the Dice metric, which is widely used in medical applications due to its focus on relative overlap.

### 2.3 RELATED WORK

In the literature on WSI segmentation in digital pathology, supervised segmentation is the standard approach for evaluating existing methods. This technique involves training an algorithm on a labeled dataset, with ground truth annotations provided by experts to guide the segmentation process (Carneiro et al. 2017; Ing et al. 2018; Komura e Ishikawa 2018). These methods often require human intervention throughout the process, particularly to refine the segmentation outputs generated by the algorithm. Ground truth annotations serve as the input, delineating one or more renal structures to be segmented in WSIs (Bel et al. 2018; Marsh et al. 2018; Gadermayr et al. 2019; Kannan et al. 2019).

Locating and detecting glomeruli in WSIs is one of the primary tasks undertaken by pathologists. Both quantitative and qualitative assessments of glomeruli are complex, time-consuming processes that requiring specialized expertise and often resulting in low inter-pathologists agreement (Haas et al. 2020). The primary goal is to provide specialists with a decision-making tool for pre-screening WSIs (e.g., glomerulus counting), thereby streamlining subsequent automated tasks, such as glomerular lesion classification. Numerous studies address the problem of glomerulus segmentation (Bel et al. 2018; Marsh et al. 2018; Gadermayr et al. 2019; Hermsen et al. 2019; Bueno et al. 2020; Altini et al. 2020; Jha et al. 2021; Jiang et al. 2021; Jayapandian et al. 2021; Bouteldja et al. 2021; Li et al. 2021; Gallego et al. 2021; Altini et al. 2023). However, most do not differentiate between classes of glomeruli, and the performance of DL models used for segmentation is often evaluated without considering the various lesions that can affect glomeruli. This oversight may lead to an overestimation of the proposed DL models' performance, especially when considering their translation to routine clinical practice.

Table 2.1 summarizes the related works selected for our study. Several insights can be drawn from Table 2.1: U-Net and its variations are prominent choices for glomerulus segmentation; the most commonly used stains are HE, PAS, and PAMS, with only three studies using TRI stain. Notably, the majority of the datasets used are private. Additionally, only one study proposes a custom DL network for glomerulus segmentation (Silva et al. 2022), while the others employ off-the-shelf DL-based segmentation models.

Lately, some studies apply DL techniques using Transformer-based models in the literature for digital histology image segmentation (Zhang e Zhang 2021; Valanarasu et al. 2021; Kawazoe et al. 2022; Saikia et al. 2023; Luo et al. 2022). The combination of Transformer-based models and U-shaped structures has become common as a robust alternative for medical image segmentation (Shamshad et al. 2023). Notably, recent works focusing on histological image segmentation with Transformer-based models have achieved satisfactory results in histology domains without glomerular structure (Zhang e Zhang 2021; Valanarasu et al. 2021; Luo et al. 2022) and with renal histology

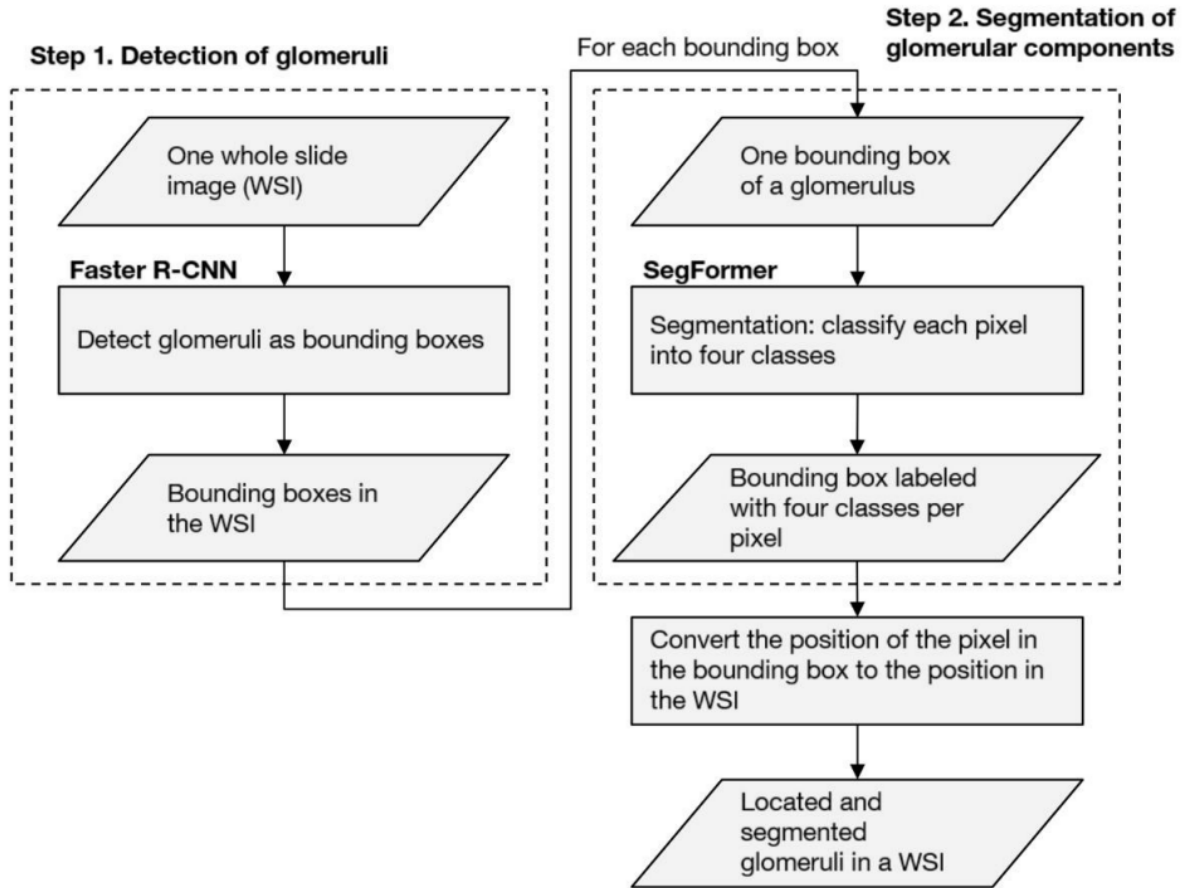
**Table 2.1** Summary of related work for glomerulus segmentation. Dashes indicate instances where information was unreported.

Specimen	Reference	#WSI	#Glom.	Method	Stain	Classes	Data set
–	(Bel et al. 2018)	15	944	FCN M-FCN U-Net	PAS	Normal Sclerosis	Private
Human	(Marsh et al. 2018)	48	3, 867	CNN(VGG-16) LoG Blob-detection Linear Regression	HE	Normal Sclerosis	Private
Mouse	(Gadermayr et al. 2019)	24	–	SW-CNN U-Net-S U-Net-D	PAS	–	Private
Human	(Hermesen et al. 2019)	60	238	U-Net	PAS	Normal Sclerosis	Private
Human	(Bueno et al. 2020)	47	1, 245	U-Net SegNet + AlexNet	PAS	Normal Sclerosis	Private
Human	(Altini et al. 2020)	26	2, 772	SegNet DeepLabv3+	PAS	Normal Sclerosis	Private
Human	(Jha et al. 2021)	61	1, 334	Mask-RCNN U-Net DeepLabv3	HE PAS PAMS	–	Private
Human	(Jiang et al. 2021)	348	8, 665	Cascade Mask R-CNN	PAS PAMS TRI	Normal Global Sclerosis Other lesions	Private
Human	(Jayapandian et al. 2021)	459	1, 196	U-Net	HE PAS PAMS TRI	–	Private
Human Mouse	(Bouteldja et al. 2021)	168	2, 611	Modified U-Net	PAS	–	Private
Human	(Li et al. 2021)	258	24, 133	U-Net	HE	Normal Sclerosis	Public
Human	(Gallego et al. 2021)	51	2, 429	U-Net	HE PAS	Normal Sclerosis	Public
Human	(Silva et al. 2022)	665	5,309	DS-FNet	HE PAS PAMS TRI	–	Public Private
Human Mouse	(Souza et al. 2023)	60	1,430	U-Net U-Net3+ Res-U-Net DeepLabV3+ MA-Net	HE PAS PAMS	–	Public Private
Human	(Altini et al. 2023)	875	23, 477	Cascade R-CNN	PAS	–	Private

with glomerular structure (Kawazoe et al. 2022; Saikia et al. 2023).

Kawazoe et al. (2022) demonstrated an automated pipeline for detecting glomeruli in 677 PAS-stained WSIs, followed by segmentation histopathological regions inside the glomeruli. The authors proposed a pipeline comprising the following two steps: (i) the detection task of glomeruli, which draws bounding boxes surrounding the glomeruli in a WSI using Faster R-CNN, and (ii) the segmentation of glomerular components, which classifies image pixels in bounding boxes labeled into five classes (Bowman’s space, glomerular tuft, crescentic, sclerotic, and background regions) using SegFormer that uses Transformer-based models, which change the position of the pixel in the bounding box to the region in the WSI, and the located and segmented glomeruli in a WSI. In Figure 2.3 shows an overview of the pipeline proposed. This pipeline, aids in medical evaluation kidney pathology by visualizing and quantifying the histopathological regions of glomerulus with the prognosis of kidney function in patients of immunoglobulin A nephropathy (IgAN).

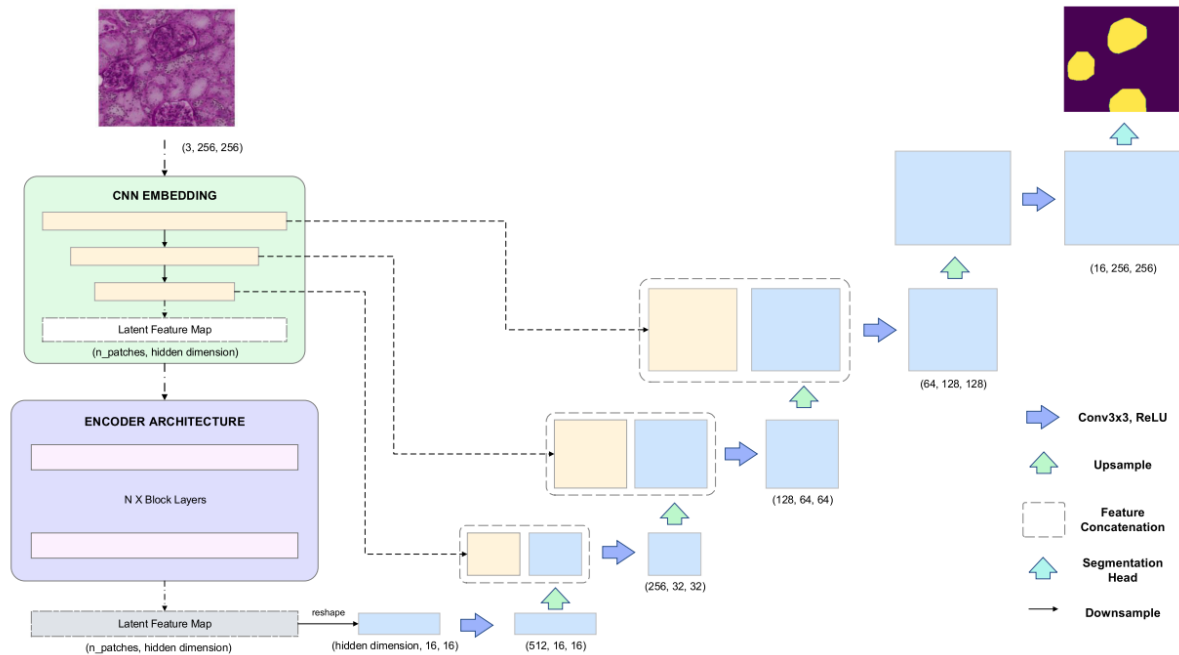
In another study, Saikia et al. (2023) proposed a new approach to segment glomeruli in 20 PAS-stained WSIs, called Multi-Layer Perceptron U-Net (MLP-U-Net), aimed at improving the diagnosis of kidney diseases. This approach incorporates both CNN-based networks and Transformer-based models. The MLP-U-Net aims to develop an embedding-encoder-decoder architecture with Multi-Layer Perceptron (MLP) encoders



**Figure 2.3** Shows an overview of the approach proposed developed in the scope of the work in (Kawazoe et al. 2022). The parallelograms represent the input or output data, while the rectangles represent the processes. Step 1: Detection of glomeruli, and Step 2: Segmentation of glomerular components into four classes: Bowman’s space, glomerular tuft, crescentic, and sclerotic. Image taken from (Kawazoe et al. 2022).

in the following Fig. 2.4. The CNN embedding module comprises downsampling CNN layers that receive input images and masks. Following the CNN embedding module, the encoder module, which utilizes Transformers and MLP, is connected to the decoder module through Convolutional and ReLU layers. Additionally, researchers conducted a comparative study of deep networks employing four approaches: U-Net, TransU-Net, Mixer-U-Net, and ResMLP-U-Net. The results indicated that MLP-based architectures provide comparable and satisfactory outcomes using the Dice metric when compared to conventional U-Net architectures.

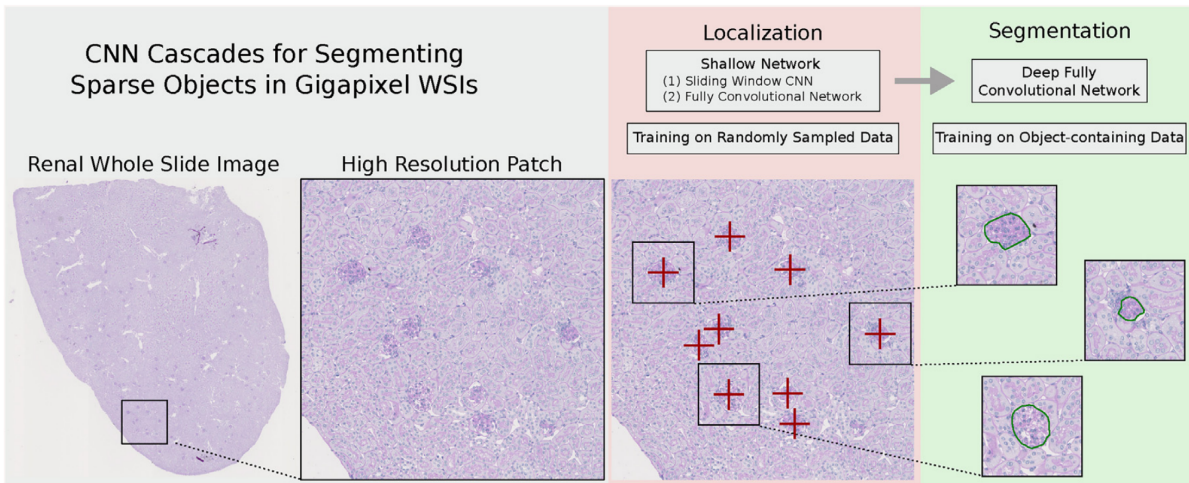
The localization and segmentation of glomeruli in mouse kidney WSIs requires capturing both broad contextual information, due to the renal biopsy process, which encompasses the entire mouse kidney. In this context, Gadermayr et al. (2019) proposed a method for glomerulus segmentation in 24 PAS-stained WSIs of mice kidney. The



**Figure 2.4** Shows an overview of the framework developed in the scope of the work in (Saikia et al. 2023). The proposed framework utilizes MLP-U-Net to generalize the encoder-decoder architecture by incorporating Multi-Layer Perceptron encoders. The CNN embedding module, which consists of downsampling CNN layers that are fed input images and masks. The encoder module comes after the CNN embedding module and is linked to the decoder module by Convolutional and ReLU layers. Image taken from (Saikia et al. 2023).

authors employed two deep neural networks: one network was adapted for glomerulus detection in WSIs, while the other was fine-tuned for segmentation based on the detection output, as illustrated in Figure 2.5. They investigated network architectures comprising the sliding window CNN method (SW-CNN) and U-Net, along with their variations. The results demonstrated that both conventional network approaches yielded satisfactory outcomes using the Dice metric. This study effectively evaluated the accuracy of automated glomerulus segmentation in mouse kidney WSIs by first proposing glomerulus localization followed by segmentation.

The identification and classification of normal and sclerosed glomeruli in WSIs using deep networks is a challenging task that requires efficient methods. To tackle this challenge, Bueno et al. (2020) conducted a study on segmenting glomeruli in 47 PAS-stained WSIs of human kidney tissue, followed by classification, as shown in Figure 2.6. The authors compared two deep networks, U-Net and SegNet, for segmenting three classes: non-glomerulus, normal glomerulus, and sclerosed glomerulus, with SegNet yielding satisfactory results. After segmentation, glomerular regions were extracted using bounding boxes, and the AlexNet classifier was applied to categorize the glomeruli as either normal or sclerosed. The combination of SegNet and AlexNet achieved satisfactory performance



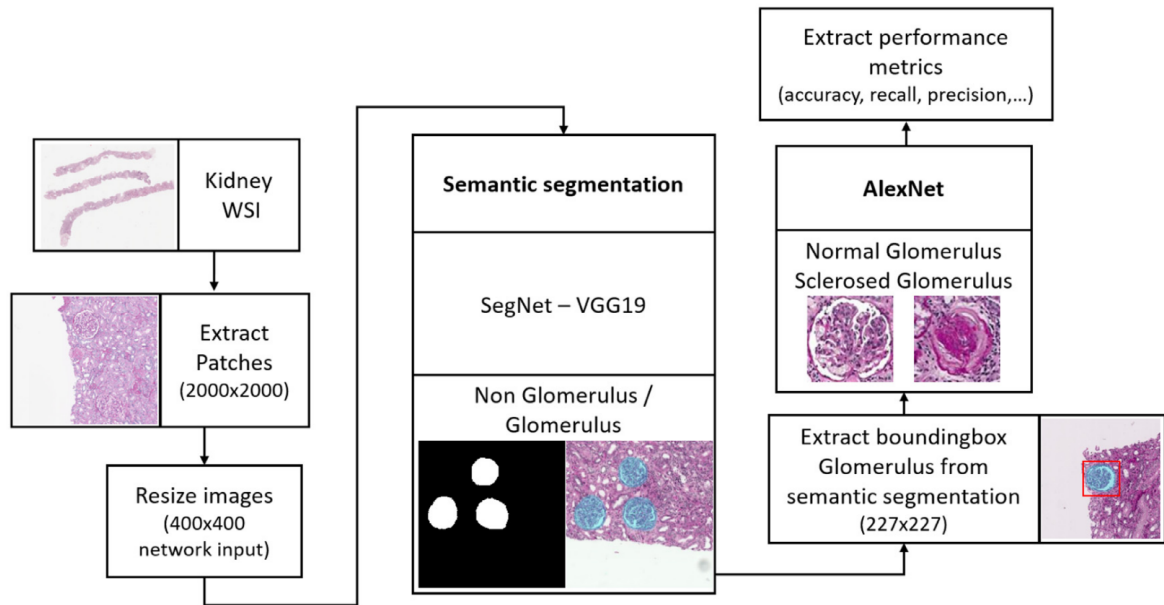
**Figure 2.5** Shows an overview of the proposed processing pipeline in the scope of the work in (Gadermayr et al. 2019). Two CNN cascades were proposed for glomerulus segmentation and compared with conventional CNN approaches. Image taken from (Gadermayr et al. 2019).

in both segmentation and classification, as measured by Accuracy and F1-Score. This study demonstrated that by applying a sequential segmentation-classification process, where segmentation of glomeruli in WSIs is performed first, the AlexNet classifier was able to reduce predictive errors between normal and sclerosed glomeruli.

In another work, Gallego et al. (2021) developed a framework for segmenting and classifying glomeruli as either normal or sclerosed in 51 WSIs of human kidney WSI stained with PAS and HE, as illustrated in Figure 2.7. The authors employed a U-Net model, which was trained on PAS-stained images and evaluated on both PAS and HE stains. The results demonstrated that the U-Net model achieved satisfactory performance using the F1-Score metric. In this study, the model was trained on WSIs from the AIDPATH dataset and tested on WSIs from four independent sites not involved in AIDPATH.

## 2.4 RELATION WITH OUR WORK

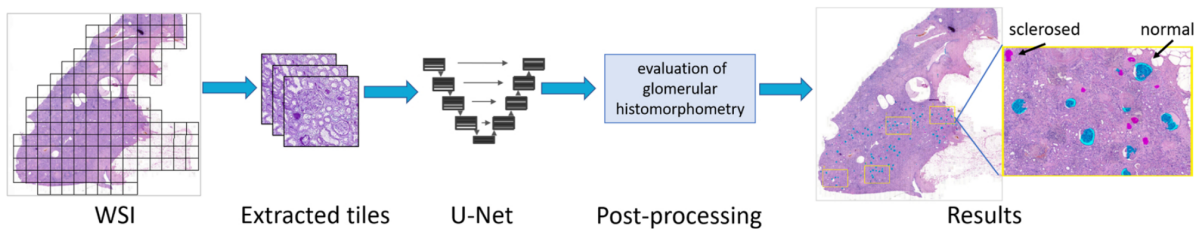
We conducted two investigations for the semantic segmentation of glomeruli. The first focused on transfer learning using supervised learning, where we trained models on a dataset of mouse images and tested them on a dataset of human images. This evaluation employed five methods: U-Net (Ronneberger et al. 2015), U-Net 3+ (Huang et al. 2020), Res-U-Net (Zhang et al. 2018), DeepLabV3+ (Chen et al. 2018), and MA-Net (Fan et al. 2020). Our study made significant strides in cross-species generalization, demonstrating that our hypothesis is indeed feasible. Experiments demonstrated that most networks, when trained on a public HE-stained mouse dataset, achieved promising results when tested on human glomerulus segmentation tasks that included HE, PAS, and PAMS stained samples. Notably, successful generalization was primarily observed when the training and test sets contained images with the same staining type; in



**Figure 2.6** Provides an overview of the workflow for the sequential segmentation and classification process, which consists of processing kidney WSIs, extracting patches, resizing images, performing semantic segmentation (SegNet), extracting bounding box, performing classification (AlexNet), and calculating performance metrics, as described in (Bueno et al. 2020). This process includes the steps of processing kidney WSIs, extracting patches, resizing images, performing semantic segmentation (using SegNet), extracting bounding boxes, conducting classification (using AlexNet), and calculating performance metrics, as outlined in (Bueno et al. 2020). The workflow is structured in two stages: (i) segmentation into two classes: Non-glomerulus (black) and glomerulus (white), followed by (ii) classification into two classes: Normal glomerulus and sclerosed glomerulus. Image taken from (Bueno et al. 2020).

cross-staining scenarios, particularly when testing on PAS and PAMS, lower and more variable scores were observed. These findings are especially relevant as they were drawn from general segmentation networks, such as those used in this study.

In the second investigation, we applied supervised learning, fine-tuning, and few-shot learning, training on a dataset of normal glomeruli images and testing on a dataset of glomeruli with segmental and global sclerosis, with a particular focus on borderless glomeruli. This evaluation explored four methods: U-Net (Ronneberger et al. 2015), DeepLabV3+ (Chen et al. 2018), SwinTransformer (Liu et al. 2021), and SegGPT (Wang et al. 2023). Our study improved the circumscription of glomeruli in scanned kidney tissue slides, representing a crucial step toward automating renal pathology diagnostics. This task becomes especially challenging when the boundaries of Bowman’s capsule, which normally outlines the glomerulus, are blurred or obscured, as seen in cases of sclerotic lesions. Sclerosed glomeruli often lack clear boundaries, complicating detection by computational models. By utilizing a foundation model with few-shot learning capabilities, this approach shows promise for enhancing segmentation accuracy



**Figure 2.7** Provides an overview of the proposed framework, as outlined in (Gallego et al. 2021). This framework is designed to segment and classify glomeruli in WSIs of human kidney specimens into normal and sclerosed categories. Image taken from (Gallego et al. 2021).

and improving diagnostic reliability in complex renal pathology cases.

U-Net, in particular, is widely recognized for its effectiveness in biomedical image segmentation and has demonstrated precise segmentation of renal histological images in WSIs (Bel et al. 2018; Gadermayr et al. 2019). Additionally, we will evaluate existing data augmentation techniques for our semantic segmenters, including methods such as flipping, rotation, Gaussian filtering, and whitening transformation, which have been well-documented in the literature.

## 2.5 CLOSURE

In this chapter, we presented the key concepts that form the foundation of our work. The next chapter will focus on the materials and methods used in our studies.



## MATERIALS AND METHODS

### Contents

---

<b>3.1</b>	<b>Data sets</b>	<b>22</b>
3.1.1	Lutnick's - Mouse	22
3.1.2	Fiocruz - Human	22
3.1.3	HuBMAP - Human	23
<b>3.2</b>	<b>Data augmentation</b>	<b>24</b>
<b>3.3</b>	<b>Semantic segmenters</b>	<b>24</b>
3.3.1	Deep-learning based non-foundation models	24
3.3.2	Foundation models	30
<b>3.4</b>	<b>Ways to improve glomerulus segmentation</b>	<b>30</b>
<b>3.5</b>	<b>Evaluation methodology of the study</b>	<b>33</b>
3.5.1	Scientific search in literature	33
3.5.2	Methods employed for segmentation	33
3.5.3	Segmentation strategy	34
<b>3.6</b>	<b>Closure</b>	<b>34</b>

---

In this chapter, we present the materials and methods used for the histopathology image segmentation task, focusing specifically on the segmentation of glomerular images. We provide an in-depth overview of dataset, segmentation models, the glomerulus segmentation problem, and evaluation methodology.

### 3.1 DATA SETS

The dataset used in this work originates from three sources: Lutnick’s (Lutnick et al. 2019), Fiocruz, and HuBMAP (Howard et al. 2020). Lutnick’s dataset has been reused with the same corpus from its original paper, containing 18 kidney sections extracted from mice specimens. The Fiocruz dataset is composed of three categories: Normal, Sclerosis, and Mixed. The Normal dataset extends the prior dataset (Souza et al. 2023), now featuring 45 WSIs of human kidney biopsies. The Sclerosis dataset is newly created and includes 37 human kidney sections, while the Mixed dataset is also new and contains 1 human kidney section. Finally, a subset of the HuBMAP dataset was selected, including 15 kidney sections from human specimens. Our experimental analyses, relied upon three datasets, whose characteristics are summarized in Table 3.1 and described next.

#### 3.1.1 Lutnick’s - Mouse

*Lutnick’s data set.* According to Lutnick et al. (2019), all histological sections were collected by following protocols approved by the Institutional Animal Care and Use Committee at the University at Buffalo, obeying the guidelines and specifications of the American Veterinary Medical Association guidelines on euthanasia. C57BL/6J mice were euthanized, and their kidneys were perfused, extracted, and embedded in paraffin. Mice were either treated with streptozotocin (STZ) to induce diabetic nephropathy or with an STZ vehicle for control. The renal biopsies were sliced from the paraffin-embedded kidney. Sections at  $2\ \mu\text{m}$  thick were stained with HE and bright-field imaged at  $0.25\ \mu\text{m}$  per pixel resolution using an Aperio ScanScope slide scanner (Leica Biosystems) in  $40\times$  magnification. This data set consists of 18 HE-stained images, each accompanied by annotations.

#### 3.1.2 Fiocruz - Human

The renal biopsies used to build the Fiocruz 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  $40\times$  magnification for the normal dataset and a Zeiss Imager.Z2 scanner with  $20\times$  magnification for the sclerosis and mixed datasets. The Fiocruz dataset contains 83 WSIs from 39 patients, annotated by our PathoSpotter group and categorized into three groups: (i) Normal, (ii) Sclerosis: Segmental and Global, and (iii) Mixed. This dataset adhered to Resolution No. 466/12 of the Brazilian National Health Council. All procedures were approved by the Ethics Committee for Research Involving Human Subjects of the Gonalo Moniz Institute, Oswaldo Cruz Foundation (CPqGM/FIOCRUZ), under Protocols No. 188/09 and No. 1817574. To ensure confidentiality, the images were anonymized, containing no identifiable patient information. These images were exclusively used for research purposes

*Normal dataset:* It comprises 45 high-resolution gigapixel images from 5 patients. Sections of  $2\ \mu\text{m}$  were stained by HE, PAS, and PAMS. The WSIs exhibit signs of no lesions

**Table 3.1** Summary of the three datasets, including descriptions of the staining types, number of WSIs, number of glomerulus, and average image dimensions used for training and testing. The staining methods explored include PAS, HE, and PAMS.

Data set	Specimen	Stain	# WSI	# Glomerulus	Average pixels (height x width)
Lutnick’s - Public	Mouse	HE	18	805	$19,511 \times 20,044$
		HE	15	221	$23,203 \times 10,639$
Fiocruz: Normal - Private	Human	PAS	15	210	$22,795 \times 11,075$
		PAMS	15	210	$22,698 \times 11,825$
		ALL	45	641	$22,899 \times 11,180$
Fiocruz: Sclerosis - Private	Human	HE	19	71	$29,211 \times 15,818$
		PAS	18	76	$26,453 \times 14,193$
		ALL	37	147	$27,869 \times 15,028$
Fiocruz: Mixed - Private	Human	HE	1	12	$26,624 \times 13,824$
HuBMAP - Public	Human	PAS	15	3,568	$29,416 \times 36,409$

and a normal appearance in the glomeruli. The selected images contain 15 WSIs stained with HE, 15 WSIs stained with PAS, and 15 WSIs stained with PAMS. 36 WSIs were used for training the segmentation models, while 9 were used for testing.

*Sclerosis dataset:* It consists of 37 high-resolution gigapixel images from 33 patients. Sections of  $2 \mu\text{m}$  thickness were stained with HE and PAS. The WSIs are derived from cases of focal segmental glomerulosclerosis (FSGS) disease. The glomeruli present can be classified into the categories of segmental or global glomerulosclerosis. The selected images include 19 WSIs stained with HE and 18 WSIs stained with PAS, which were used only to test the segmentation model.

*Mixed dataset:* It contains a single high-resolution gigapixel image from 1 patient. Sections of  $2 \mu\text{m}$  thickness were stained with HE. The WSI is categorized into three classes: normal, segmental sclerosis, and global sclerosis. This WSI was exclusively used for training via fine-tuning and for querying through SegGPT of the segmentation model. For the per-crop evaluation, the same WSI was utilized, consisting of only 12 glomerulus patches extracted from the WSI, with 4 patches from each class: normal, segmental sclerosis, and global sclerosis.

### 3.1.3 HuBMAP - Human

Sponsored by the National Institutes of Health (NIH), the Human BioMolecular Atlas Program (HuBMAP) managed a glomerulus segmentation competition in the Kaggle(Howard et al. 2020) platform. This challenge involved developing a supervised model for glomerulus segmentation using a set of 20 human kidney WSIs. The dataset includes 11 fresh frozen and 9 formalin-fixed paraffin-embedded (FFPE) PAS-stained kidney images. Of these WSIs, 15 are designated for training, and 5 are reserved for testing. Test images and annotations are not available, and model results can only be verified through submissions on the Kaggle platform. The HuBMAP dataset also contains ad-

ditional information, including anonymized patient data, for each image. These images were used exclusively for the HuBMAP “Hacking the Kidney” competition.

## 3.2 DATA AUGMENTATION

The process of increasing the quantity and diversity of the dataset without the need to collect new data is known as data augmentation. It relies on existing datasets to make small transformations, generating modified copies, or in other words, new data samples. As data augmentation strategies, we applied resizing, vertical and horizontal flip, random rotation in intervals of 90 degrees, shift scale rotation, Gaussian noise, Gaussian blur, random brightness contrast, and random hue saturation value. These techniques were highlighted in the articles from Table 2.1 and in the Kaggle (Howard et al. 2020) competition focused for glomeruli segmentation. Henceforth, we continued with these strategies in our two studies using deep networks.

## 3.3 SEMANTIC SEGMENTERS

Semantic segmentation is the task of categorizing each pixel of an image into distinct regions. In this segmentation, each region is associated with a label corresponding to a semantic class.

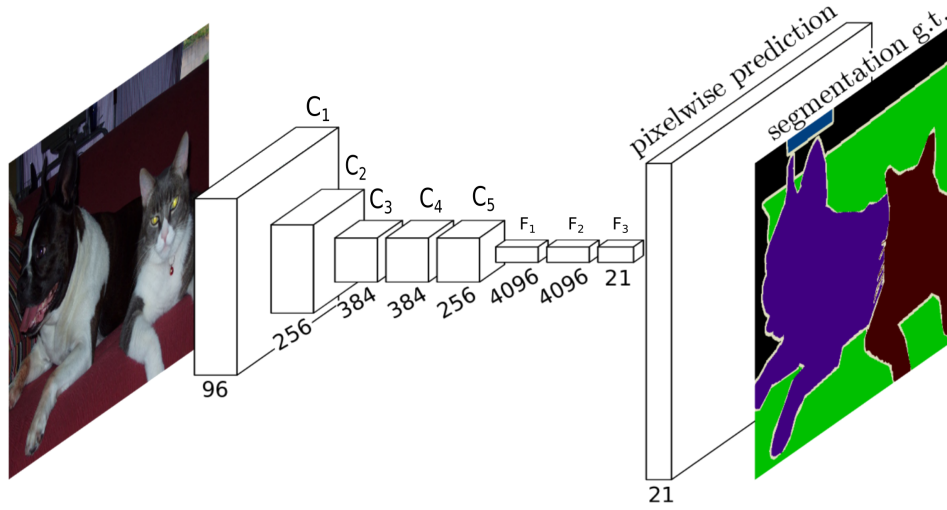
*Region-based* This approach tries to solve the problem of finding the boundaries between regions of the image. Regional-based methods, such as region-growing (Adams e Bischof 1994), explore the discontinuities between pixel intensities to find these boundaries.

This approach is capable of accurately segmenting regions that share the same properties and are spatially separated, generating interconnected regions (Bankman 2000). Some object segmentations, manual intervention is required to determine the initial seed placement point. Therefore, it is necessary to place a seed for each region that needs to be segmented (Gonzalez 2009).

### 3.3.1 Deep-learning based non-foundation models

As demonstrated across multiple research domains, deep learning has been successful in computer vision (CV) in solving complex tasks (Chai et al. 2021). In the case of image segmentation, convolutional networks are the main approach. These networks comprise a series of convolutional filters, which the parameters can be learned from the data. The following discussion focuses on FCN (Long et al. 2015), U-Net (Ronneberger et al. 2015), U-Net 3+ (Huang et al. 2020), Res-U-Net (Zhang et al. 2018), DeepLabV3+ (Chen et al. 2018), MA-Net (Fan et al. 2020), and SwinTransformer + U-Net (Liu et al. 2021; Cao et al. 2022), which are the main models of deep networks used in segmentation tasks.

*FCN* This method is an adaptation of the Visual Geometry Group (VGG) network (Simonyan e Zisserman 2014) for the semantic segmentation task. fully convolutional

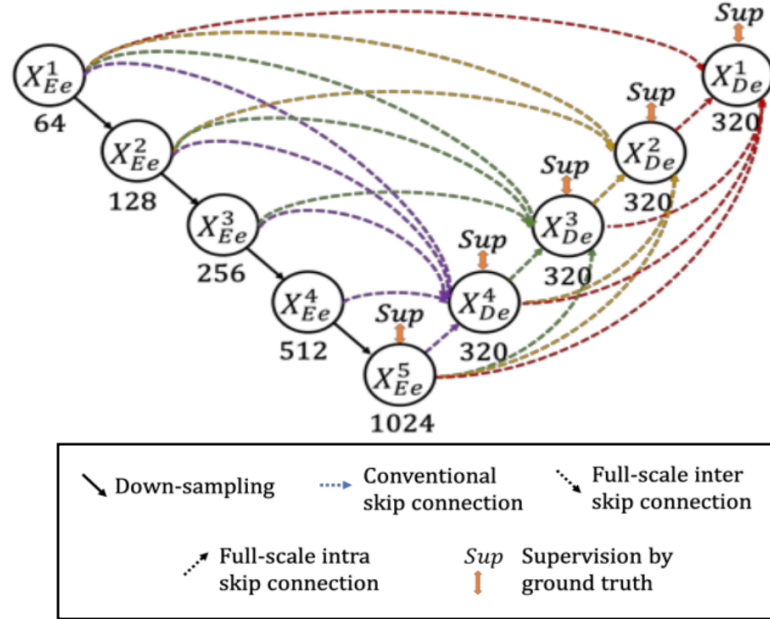


**Figure 3.1** FCN developed in the scope of the work in (Long et al. 2015), adapted from a VGG classification network (Simonyan e Zisserman 2014). The blocks  $C_1$ ,  $C_2$ ,  $C_3$ ,  $C_4$ ,  $C_5$  represent the convolutional stages just as they are present in the VGG network. The blocks  $F_1$ ,  $F_2$ ,  $F_3$  represent the fully connected layers that have been replaced by convolutional layer with a kernel of  $1 \times 1$ . The block named pixelwise prediction represents a upsampling layer used to return the convoluted image to the size of the output matrix. Image taken from (Long et al. 2015).

network (FCN) was proposed by (Long et al. 2015) as the first segmentation network capable of being trained end-to-end for pixel-wise prediction. This indicates that, for each iteration of the deep network’s training, the model maps the input image to the desired outputs using a certain amount of labeled data. In other words, the model learns to extract useful features from the data and to use these features for making predictions without passing through a pipeline of techniques. The network leverages the spatial correspondence and translational invariance present in convolutional networks. In a classification network, such as VGG, convolution and pooling operations maintain a certain degree of spatial correspondence between the input image and the output features. To ensure this equivalence in the FCN, the authors converted the fully connected layers to 2D convolutions with a  $1 \times 1$  kernel. Figure 3.1 illustrates the fundamental architecture of the FCN, where the network learns to generate dense predictions for semantic segmentation task.

*U-Net* is one of the first networks for medical image segmentation based on an encoder-decoder architecture, forming a U shape. This architecture is composed by a reduction step (encoder) and an expansion step (decoder). The encoder reduces spatial dimensions while increasing channels, whereas the decoder increases spatial dimensions and reduces channels, ultimately restoring spatial dimensions to make pixel-wise predictions for the input image. The encoder can be based on a standard deep-learning classification network, such as FCN, where each block in the downsampling operation consists of convolution and max pooling operations. The upsampling stage utilizes successive blocks to combine





**Figure 3.3** The U-Net 3+ architecture, developed in (Huang et al. 2020). U-Net 3+ integrates multi-scale features through redesigned skip connections and full-scale deep supervision, producing a segmentation map that is both position-aware and boundary-enhanced. Image adapted from (Huang et al. 2020).

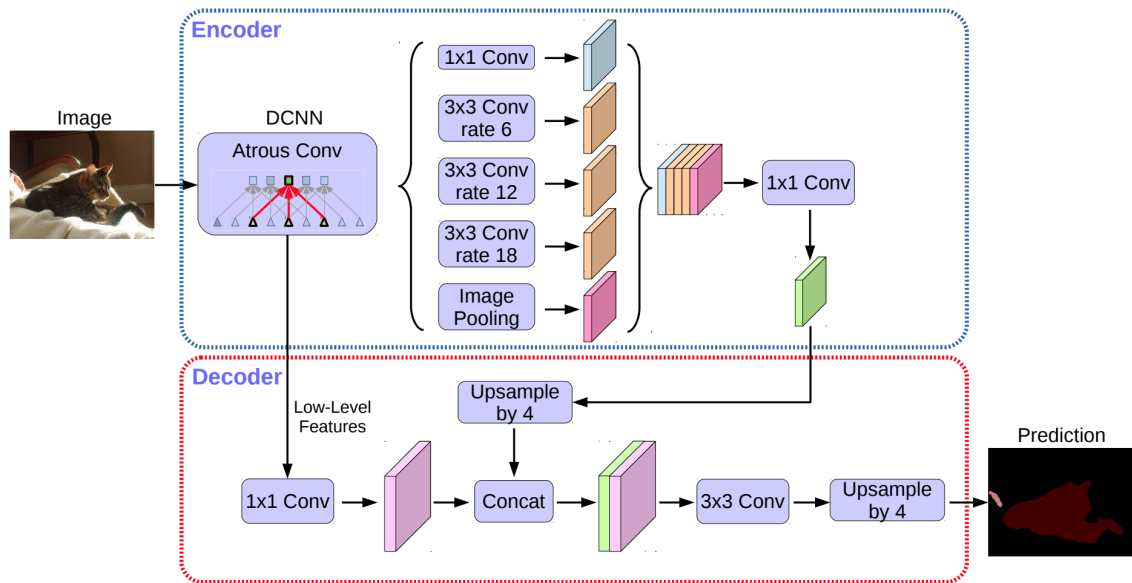
and decoder blocks. This architecture improves information flow through skip connections within residual units and between encoding and decoding layers, helping to prevent gradient degradation and requiring fewer parameters, as shown in Figure 3.4. By combining the strengths of residual networks with U-Net’s segmentation capabilities, Res-U-Net achieves competitive or superior performance in semantic segmentation tasks with improved efficiency.

*DeepLabV3+* is a deep learning architecture designed for semantic image segmentation (Chen et al. 2018), created from improvements on the DeepLabV3 network (Chen et al. 2017), by employing an encoder-decoder structure. This architecture uses the encoder stage and adds a simple yet effective decoder stage to refine the segmentation results especially along object boundaries. Even though the encoder can process multi-scale contextual information by applying dilated convolution at multiple scales, DeepLabV3+ was not properly designed to obtain long-range dependencies in the deep learning process. Figure 3.5 illustrates the DeepLabV3+ architecture (Chen et al. 2018).

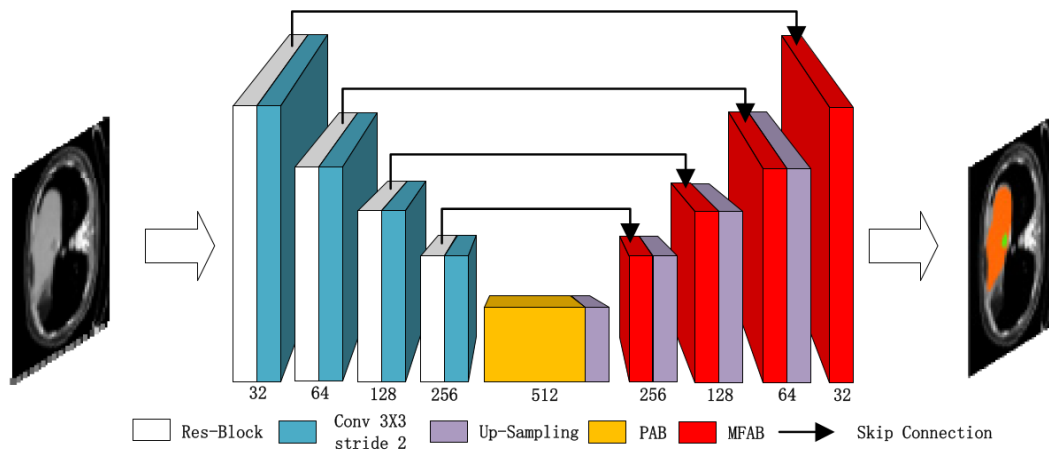
*MA-Net* is the multi-scale attention network built upon the U-Net architecture as its basic structure, incorporating a residual mechanism in the encoder to enhance feature extraction (Fan et al. 2020). Additionally, the convolutional block attention module (CBAM) is divided into channel attention and spatial attention modules, which are







**Figure 3.5** The DeepLabV3+ encoder-decoder architecture, developed in (Chen et al. 2018). The encoder step encodes multi-scale contextual information by applying atrous convolution at multiple scales, while the simple yet effective decoder step refines the segmentation results along object boundaries. Image taken from (Chen et al. 2018).



**Figure 3.6** The architecture of MA-Net network, developed in (Fan et al. 2020). The encoder step includes Res-blocks, consisting of convolution layers and residual connections, to extract high-dimensional feature information, along with the PAB, while the decoder step employs the MFAB to capture channel dependencies, enriching multi-scale semantic information and enhancing network performance. Image taken from (Fan et al. 2020).

effectively combine them with high-level features, enhancing segmentation accuracy and contextual understanding. The MA-Net is a deep learning architecture that consists of

a self-attention mechanism for adaptive feature extraction using two stages: (i) position-wise attention block (PAB), which covers feature inter-dependencies between pixels in spatial dimensions and (ii) a multi-scale fusion attention block (MFAB), which captures the channel dependencies between any feature map by multi-scale semantic feature fusion. Figure 3.6 illustrates the MA-Net architecture.

Swin Transformer + U-Net is a Transformer-based architecture for medical image segmentation that combines a U-Net-like encoder-decoder structure with the Swin Transformer (Liu et al. 2021; Cao et al. 2022). It tokenizes image patches and processes them through a hierarchical Swin Transformer encoder with shifted windows, capturing both local and global semantic features. The decoder, constructed from symmetric Swin Transformer blocks, uses patch-expanding layers to restore spatial resolution through up-sampling. This design enables Swin Transformer + U-Net to efficiently model long-range dependencies and hierarchical features. Figure 3.7 depicts the architecture of Swin Transformer + U-Net.

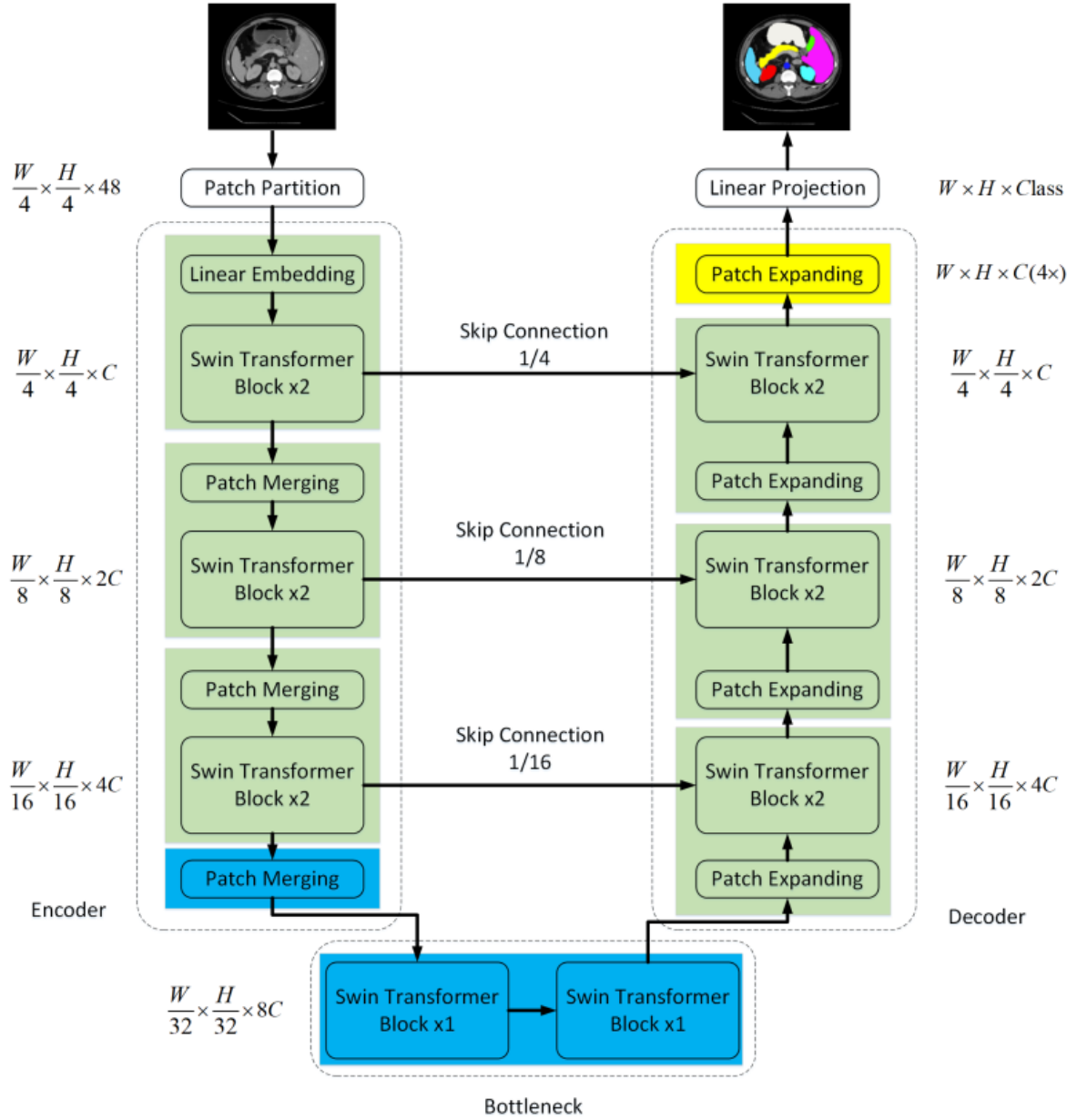
### 3.3.2 Foundation models

Image segmentation is a long-standing challenge in the field of CV, having been the focus of extensive research for many years, as demonstrated by both traditional non-deep learning methods and deep learning approaches. With the advent of foundation models (FMs), methods aimed at image segmentation have entered a new era, leveraging the capabilities of these advanced models. The term “foundation model” refers to machine learning (ML) models that are trained on large-scale, generalized datasets, enabling them to perform a wide variety of tasks across different domains (Zhou et al. 2024). These approaches not only provide superior segmentation performance but also introduce new segmentation capabilities that have not been seen before in the context of deep learning. Furthermore, it is faster and more cost-effective to utilize pre-trained FMs to develop new ML applications, rather than training bespoke ML models from scratch.

*SegGPT* is a generalist model developed for context-aware segmentation tasks, utilizing a FM that distinguishes it from non-foundation models (Wang et al. 2023). It approaches training as an in-context coloring task, applying random color mapping to each data sample. We adapted this model for gigapixel histopathological images of glomeruli, allowing it to function without a conventional training phase and thus avoiding the need for extensive task-specific training or fine-tuning. Instead, SegGPT leverages feature representations from a few query samples during inference. Figure 3.8 illustrates the SegGPT FM (Wang et al. 2023).

## 3.4 WAYS TO IMPROVE GLOMERULUS SEGMENTATION

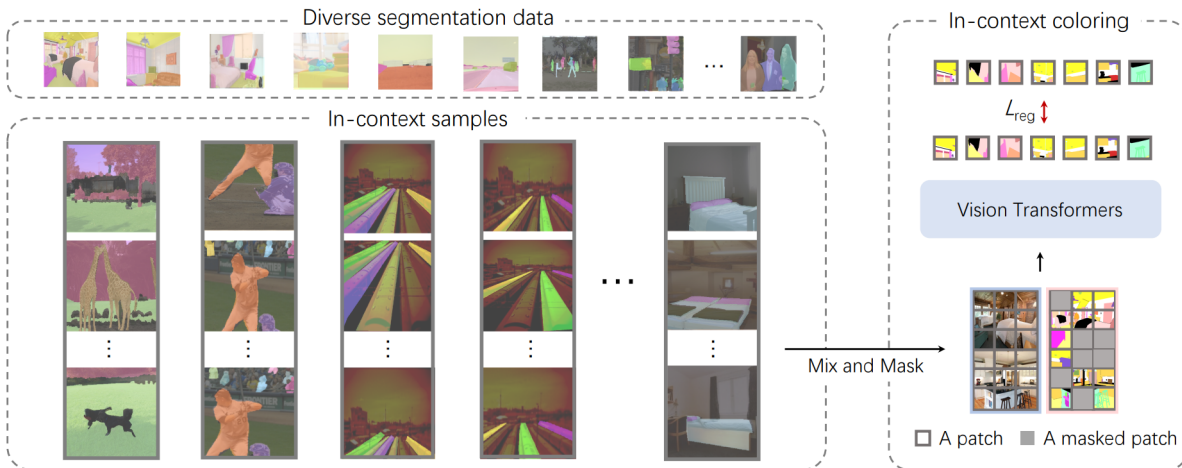
Our study focuses on the glomerulus, a roughly spherical network of tiny blood capillaries responsible for blood filtration. Given its critical function, the localization and segmentation of glomeruli provide valuable information that pathologists often extract from kidney WSI. However, locating and segmenting glomeruli can be time-consuming and prone to



**Figure 3.7** The architecture of SwinTransformer + U-Net developed in (Cao et al. 2022), which is composed of encoder, bottleneck, decoder and skip connections. Encoder, bottleneck and decoder are all constructed based on swin transformer block (Liu et al. 2021). Image taken from (Cao et al. 2022).

errors. To address that, there is a growing interest in developing automated glomerular segmentation approaches that offer fast and reliable support for pathologists’ decision-making processes. The following discussion highlights four alternatives to enhance the segmentation task with limited training data: (i) data augmentation, (ii) improvement of deep learning methods, (iii) the emergence of foundation models, and (iv) refinement of the dataset.

In the first alternative, data augmentation refers to techniques used to artificially increase the dataset (see Section 3.2). In summary, the main benefits of this technique include reducing overfitting, increasing model accuracy, and assisting in scenarios with limited data sources. In the second alternative, we discuss the improvement of deep learning methods through the use of transfer learning techniques to leverage pre-trained models from rodents and fine-tune them on human data. This approach is outlined in our publication titled “Mouse-to-Human Transfer Learning for Glomerulus Segmentation” (Souza et al. 2023) (see Chapter 4). In the third alternative, we explore the use of the foundation model SegGPT to assess borderless glomeruli, as discussed in the submission titled “The Problem of Segmenting Global Glomerulosclerosis in Gigapixel Histopathological Images: The Borderless Glomeruli” (see Chapter 5), which could be crucial for the diagnosis of certain kidney diseases. Lastly, collaboration with renal pathologists is essential to ensure that the dataset is representative and well-labeled, as well as to expand the dataset to make it more diverse for the segmentation task. In our group, PathoSpotter, we have a suite of intelligent tools to support pathologists’ daily



**Figure 3.8** Illustration of overall training framework of SegGPT, developed in (Wang et al. 2023). The authors incorporate diverse segmentation data, including part, semantic, instance, panoptic, person, medical image, and aerial image segmentation, and transform them into the same format of images. They adopt a general Painter (Wang et al. 2023) framework with in-context coloring as the training objective and a random coloring scheme for more flexible and generalizable training. Image taken from (Wang et al. 2023).

practice <sup>1</sup>.

## 3.5 EVALUATION METHODOLOGY OF THE STUDY

### 3.5.1 Scientific search in literature

This study was based on a search for scientific articles in the following datasets: Scopus<sup>2</sup>, IEEE<sup>3</sup> (*Institute of Electrical and Electronics Engineers*), *Engineering Village*<sup>4</sup>, CAPES Journal Portal<sup>5</sup>, *Google Scholar*<sup>6</sup>, and the gold-standard repository for medical journals, PubMed<sup>7</sup>. In these databases, we reviewed articles containing the following keywords: *glomerulus*, *glomerulosclerosis*, *glomerulus segmentation*, *kidney biopsies*, *kidney histopathological images*, *deep-learning e foundation model*. The criteria for selecting these works were threefold: (i) they perform glomerulus segmentation, (ii) they utilize whole slide images (WSI), even if processing patches, and (iii) the articles based on deep learning methods on the topic began to be published in 2018 and continue to the present day.

### 3.5.2 Methods employed for segmentation

Instead of using just one model for each network architecture in the two studies focused on glomeruli, we first evaluated the five most relevant methods from the literature to investigate mouse-to-human transfer learning. In the second study, we applied three of the most relevant methods, both with and without fine-tuning, along with a new approach from the literature utilizing a foundation model.

In our first study to evaluate the similarity between mice and humans, we used transfer learning to investigate whether five deep learning models: U-Net, U-Net 3+, Res-U-Net, DeepLabv3+, MA-Net —trained on mice data can segment glomeruli using human data. In our second study, we are investigating the problem of segmenting global glomerulosclerosis from a human dataset. The initial setup involved loading three deep learning methods: U-Net, U-Net 3+, and Swin-Transformer + U-Net, with weights pre-trained on the HuBMAP dataset (Howard et al. 2020). We then performed training with these three networks, both with and without fine-tuning, and finally, we conducted a context-aware analysis without training using the foundation model: Seg-GPT.

The number of images utilized in the training and testing subsets, as well as the results obtained using the Dice metric, is detailed in Chapters 4 and 5, for the first and second studies, respectively. There is a variation in the number of WSIs from the Fiocruz dataset used for training and testing in the first and second studies, as the first study had a limited number of annotated human sections.

---

<sup>1</sup><https://pathospotter.bahia.fiocruz.br/>

<sup>2</sup><http://www.scopus.com/>

<sup>3</sup><http://ieeexplore.ieee.org/Xplore/home.jsp>

<sup>4</sup><http://www.engineeringvillage.com/>

<sup>5</sup><http://www-periodicos-capes-gov-br.ez10.periodicos.capes.gov.br/>

<sup>6</sup><https://scholar.google.com.br/>

<sup>7</sup><https://pubmed.ncbi.nlm.nih.gov>

### **3.5.3 Segmentation strategy**

Our segmentation strategy begins with the automatic division of the image, which is cropped into 1,024 x 1,024 pixels, with a padding size of 256 pixels. To manage the high memory process and enhance training speed, each patch is resized to 320 x 320 pixels before being used as input for the segmentation models. The results from each model for the corresponding patches are summed, and a sigmoid function is then applied to the resulting image. Finally, the masks containing the segmentation of the glomerulus are resized to match the original input size, and the resulting masks are evaluated using a Dice metric.

## **3.6 CLOSURE**

In this chapter, we presented the materials and methods important that ground our work. In the following chapters, we will discuss the details of our study within the context of glomerulus segmentation, starting with mouse-to-human transfer learning.

# TRANSFER LEARNING USING CROSS-SPECIES DATA FOR NORMAL GLOMERULUS SEGMENTATION

## Contents

---

<b>4.1</b>	<b>Background . . . . .</b>	<b>35</b>
<b>4.2</b>	<b>Mouse-to-human transfer learning . . . . .</b>	<b>36</b>
<b>4.3</b>	<b>Materials and methods . . . . .</b>	<b>38</b>
	4.3.1 Datasets . . . . .	40
	4.3.2 Methods . . . . .	40
	4.3.3 Implementation details . . . . .	41
<b>4.4</b>	<b>Results and discussion . . . . .</b>	<b>41</b>
	4.4.1 Quantitative analysis . . . . .	41
	4.4.2 Qualitative analysis . . . . .	44
<b>4.5</b>	<b>Closure . . . . .</b>	<b>48</b>

---

In this chapter, we present our first study focused on the task of segmentation, specifically addressing mouse-to-human transfer learning for glomerulus segmentation. We provide an overview of the study, including implementation details, experimental analysis, and the corresponding results and discussion.

## 4.1 BACKGROUND

It is usual to conduct animal experiments before clinical tests on humans while performing scientific medical studies. Some of these studies include histological analysis of representative tissue samples, which must be collected from specific animals and under unique circumstances to generate findings that can apply to humans. Among several animals, mice are frequently selected for these studies, not only for their genetic and physiological similarities with humans (Kim et al. 2007) but also because

the rodents are inexpensive to breed and their tiny sizes take less space in the laboratory facilities (Smith e Corrow 2005). That preemptive mouse experimentation also benefits the study and diagnosis of kidney diseases inside a field called nephropathology, which is ultimately the domain of this study. Kidney biopsies may be required for diagnosing renal diseases depending on the problem’s complexity. As a common approach, these renal samples are processed and scanned into WSI, which are easier to be shared among specialists while analyzed on computational environments (Barisoni et al. 2013; Farris et al. 2017; Santos et al. 2019; Chagas et al. 2022).

DL techniques have gained prominence in the literature for addressing automatic segmentation tasks, consistently achieving state of the art results across various domains, including medical imaging and, more specifically, WSI segmentation (Bel et al. 2018; Gadermayr et al. 2019; Ginley et al. 2020; Jiang et al. 2021; Silva et al. 2022). However, DL models often require large annotated datasets for effective training, either from scratch or through fine tuning on target domains. Given the limited availability of annotated human WSIs, this study explores the potential of leveraging transfer learning by using models trained on mouse WSIs to segment human glomeruli.

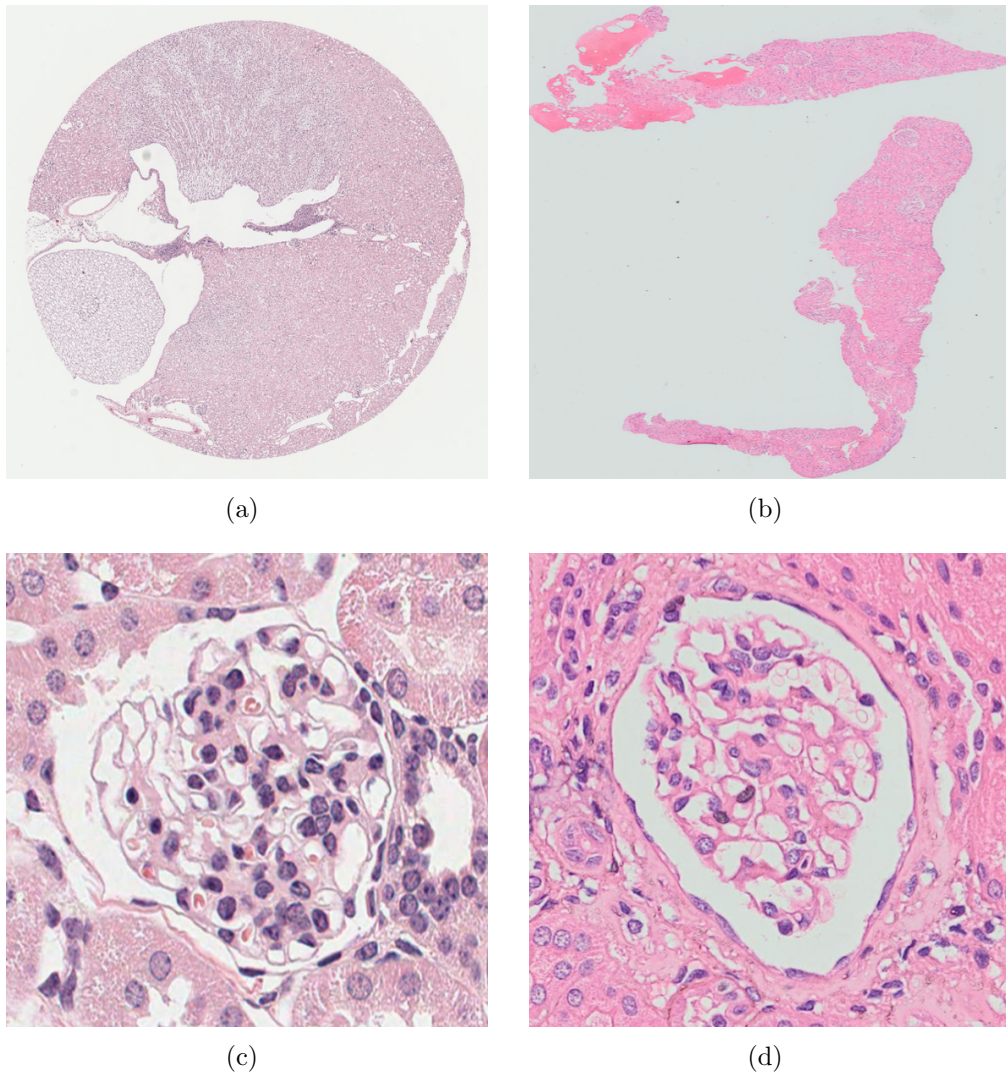
We hypothesize that, due to clinically known similarities between human and mouse glomeruli, it is possible to achieve accurate glomerular segmentation on human samples using models trained exclusively on mouse data. Among the various kidney structures, this work focuses on the glomerulus, a spherical network of capillaries responsible for blood filtration. Detecting and segmenting glomeruli in WSIs is crucial, as this information supports diagnostic workflows. However, manual glomerulus annotation is time consuming and prone to error, making automated segmentation a promising solution to assist pathologists by offering a fast and reliable tool for decision making.

## 4.2 MOUSE-TO-HUMAN TRANSFER LEARNING

The challenge of collecting large volumes of annotated data, particularly from human biopsies, remains a significant barrier. This constraint naturally raises the following research question: Is it possible to accurately segment human glomeruli in WSIs using a model trained solely on mouse glomerular images? This study aims to address this question by examining cross species transfer learning and evaluating its effectiveness for glomerulus segmentation.

From the histology perspective, it is worth noting that there is no obstacle to mixing mouse and human samples. The renal histological structures are similar across these species, as illustrated in Figure 4.1, despite variations in size. Mice are widely used in the study of various human diseases, including infectious, autoimmune, metabolic disorders, and drug toxicity. Thus, a system capable of identifying and correlating pathological changes in kidney WSIs for both humans and mice would be highly valuable for research and therapeutic development. Although there are some works using mouse and human species outside the domain of histology (Chater et al. 2021; Hossain et al. 2021), few deep-learning-based studies have proposed exploring images of histology samples of these species, and only in isolated (same specie for training and testing) evaluation protocols (Bouteldja et al. 2021; Simon et al. 2018; Ginley et al. 2019; Lutnick et al. 2019).





**Figure 4.1** Examples of two renal WSI and two glomeruli: (a) A mouse WSI; (b) A human WSI; (c) A mouse glomerulus image from Lutnick’s dataset; (d) A human glomerulus image from the Fiocruz dataset.

Bouteldja et al. (2021) developed a custom U-Net network for automated multi-class segmentation of glomerular images of different mammalian species, not only mice and humans. Simon et al. (2018) modified a feature extractor based on local binary pattern (LBP) to feed a support vector machine (SVM) model for glomerulus detection in human and mouse WSIs, separately. Ginley et al. (2019) proposed a deep-learning-based approach to quantify the number of nuclei, capillary lumens, and Bowman spaces from histology images of humans and mice; a set of features is used to describe the structural progression of diabetic nephropathy through a recurrent neural network used; yet, a DeepLabV2 network is used for glomerular segmentation on WSIs. Lutnick et al. (2019) applied a DeepLabV2 network to segment glomerulus and internal glomerular structures

from human and mouse renal tissue slides.

Given the DL dominance in computer vision and segmentation tasks, instead of evaluating human or mouse glomerular segmentation in an isolated fashion, we propose an evaluation of DL segmentation models for the cross-species generalization task, *i.e.*, to use a model trained with mouse data to segment human glomerulus.

**Contributions:** Considering the similarities between mice and humans, the main contribution of our work is the investigation of the cross-species compatibility of human and mouse data for glomerulus segmentation. Our main goal is to investigate if a DL model, trained on mouse data, can segment the human glomerulus. In addition to bringing experiments that aim to demonstrate our initial hypothesis, we also show a qualitative analysis of a data set of glomeruli with unbounded shapes. This last study aims to present the difficulty of the segmentation task when trying to predict on top of bruised-boundary glomeruli.

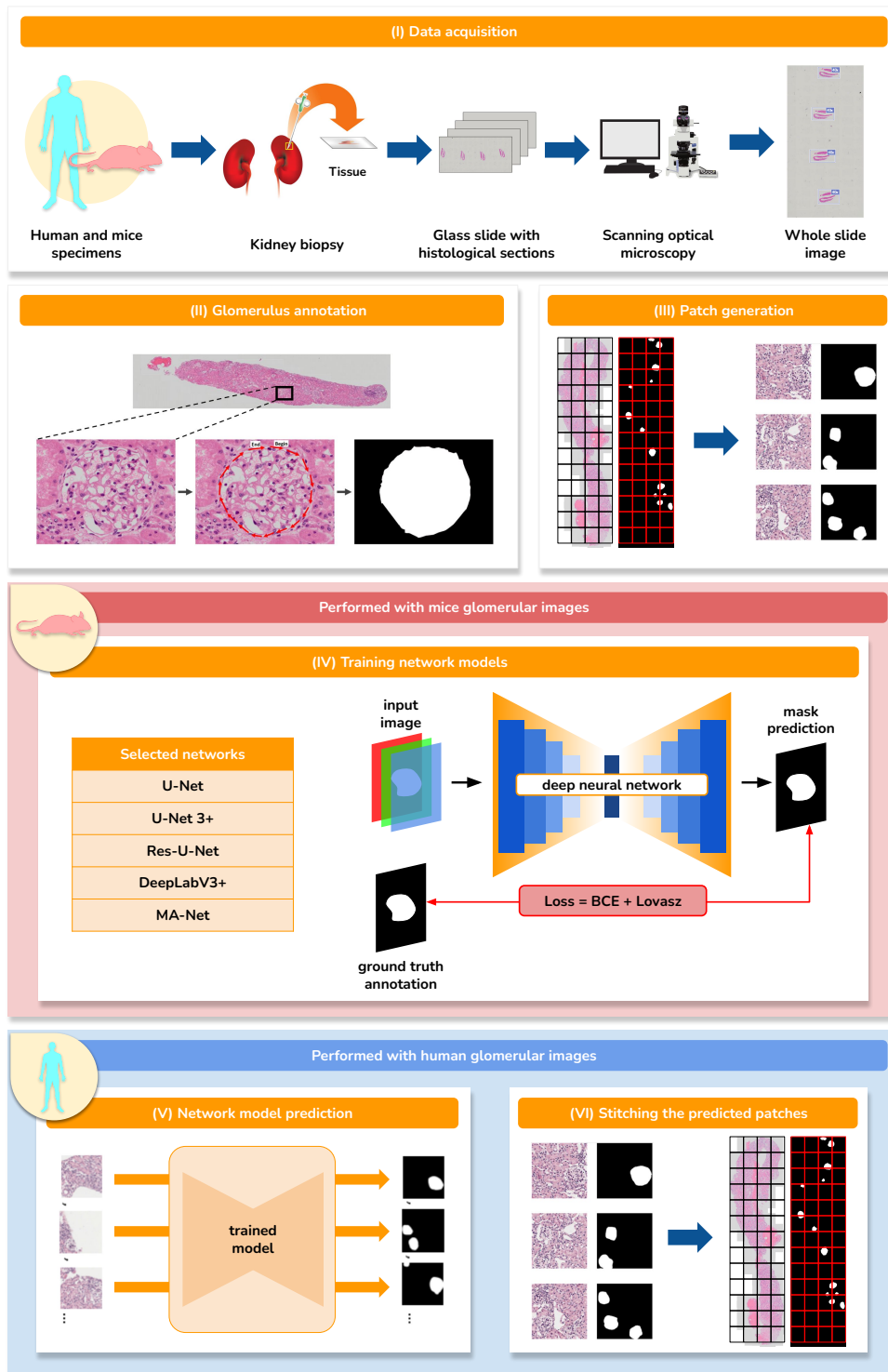
### 4.3 MATERIALS AND METHODS

In this section, we provide a comprehensive overview of the materials and methodologies utilized in our study to investigate the segmentation of human glomeruli using models trained on mouse WSIs. This includes a detailed description of the datasets, methods, and implementation strategies employed throughout the research.

The analytical protocol shown in Fig. 4.2 illustrates the outline of the proposed study split as follows: (i) data acquisition, (ii) glomerulus annotation, (iii) patch generation, (iv) training of the network models with mice glomerular images, (v) network model prediction on human samples, and finally (vi) stitching the predicted patches of human glomeruli. **The first step** consists of the extraction of kidney biopsy sections from mice and humans with  $40\times$  magnification, followed by staining with HE, PAS, and PAMS. **In the second step**, the Cytomine<sup>1</sup> software was used to perform manual annotation of the glomeruli to support the training of the network models. **At the third step**, the WSIs were divided into patches of  $1024\times 1024$  pixels with padding size of 256 pixels; each patch was resized to  $320\times 320$  pixels due to high-memory footprint to increase training speed. **In the fourth step**, the generated patches from mice samples are used to train the following networks: U-Net (Ronneberger et al. 2015), U-Net 3+ (Huang et al. 2020), Res-U-Net (Zhang et al. 2018), DeepLabV3+ (Chen et al. 2018), and MA-Net (Fan et al. 2020). We highlight that all networks had the EfficientNet-b1 (Tan e Le 2019) architecture as the backbone (encoder). The next steps refer to the segmentation of human glomerular samples using the models trained on mice glomerular images. **In the fifth step**, the best prediction model was selected for each architecture, and the segmentation performance was assessed through the Dice score. Finally, **in the sixth step**, the resulting predicted patches from all models are stitched to compound the resulting semantic segmentation masks.

---

<sup>1</sup><https://cytomine.com/>



**Figure 4.2** Analytical protocol for this study. From (i) to (vi): (i) kidney biopsy samples are collected from human and mouse specimens, whose histological sections are processed, stained and scanned into WSIs; (ii) glomeruli are manually annotated from sections of kidney biopsies; (iii) each section is divided into patches; (iv) five types of architectures were trained with mice glomerular images; (v) the best model of each architecture is selected to predict human glomerulus classes in each patch; (vi) finally, the resulting masks are stitched to compose the final segmented WSI mask.

### 4.3.1 Datasets

The datasets used in this study include one introduced by (Lutnick et al. 2019) (See Section 3.1.1), which focuses on mice, and our own collection for humans, referred to as Fiocruz (Souza et al. 2023) (See Section 3.1.2 - Normal dataset). Lutnick’s dataset has been reused in its original form, containing 18 kidney sections extracted from mouse specimens. The Fiocruz dataset is an extension of the previously published Fiocruz dataset (Souza et al. 2022) and now comprises 42 kidney sections collected from human specimens. Table 4.1 provides a summary of the details for both datasets.

### 4.3.2 Methods

Segmenting semantic labels requires pinpointing the contour of structures and pixel-level accuracy. The semantic network models commonly have an encoder-decoder architecture: the encoder learns hierarchical features while it downsamples the input image resolution; the decoder receives the aggregated encoded features and generates the segmentation mask. In this present work, we selected five types of segmentation networks: U-Net (Ronneberger et al. 2015), U-Net 3+ (Huang et al. 2020), Res-U-Net (Zhang et al. 2018), DeepLabV3+ (Chen et al. 2018), and MA-Net (Fan et al. 2020). Each network relies on a different deep learning segmentation paradigm, as further detailed.

**U-Net based networks:** The U-Net is a typical encoder-decoder architecture, where each encoding layer is connected to each respective upscaling layer of the decoder, forming a U shape. Different U-Net variations can be found for medical image segmentation tasks, such as U-Net (Ronneberger et al. 2015), U-Net 3+ (Huang et al. 2020), and Res-U-Net (Zhang et al. 2018). U-Net 3+ combines the multi-scale features by re-designing skip connections and capturing feature maps in full scales for accurate segmentation with fewer parameters. In the U-Net 3+ network, each decoder layer bridges smaller- and same-scale activation maps from the encoder and larger-scale activation maps from the decoder. The Res-U-Net is a similar U-Net variation that also brings the strengths of residual learning; this architecture uses residual units with identity mapping instead of encoder and decoder blocks from the classic U-Net architecture. The main goal of this approach is to improve skip connections inside the residual units and between the encoding and decoding layers, ultimately facilitating information propagation with fewer parameters.

**DeepLabV3+ network:** The DeepLabV3+ is a deep learning architecture designed for semantic image segmentation, created from improvements on the DeepLabV3 network

**Table 4.1** Summary of the data sets. Lutnick’s data set comprises only mouse data, while Fiocruz is composed of human data.

Data set	Stain	# WSI		# Glomerulus		Average pixels (height × width)
		Train	Test	Train	Test	
Lutnick’s	HE	14	4	634	171	19,511 × 20,044
	HE	6	6	103	102	23,111 × 10,178
Fiocruz	PAS	9	6	149	61	22,251 × 10,441
	PAMS	9	6	133	77	21,881 × 11,214

(Chen et al. 2017). Even though the encoder can process multi-scale contextual information by applying dilated convolution at multiple scales, DeepLabV3+ was not properly designed to obtain long-range dependencies in the deep learning process.

**MA-Net network:** The multi-scale attention network is an architecture that consists of a self-attention mechanism for adaptive feature extraction using two stages: (i) PAB, which covers feature inter-dependencies between pixels in spatial dimensions and (ii) a MFAB, which captures the channel dependencies between any feature map by multi-scale semantic feature fusion.

### 4.3.3 Implementation details

Based on the survey presented in Table 2.1 (see Chapter 2), the initial experiments in our work, and glomeruli segmentation in the Kaggle competition (Howard et al. 2020), we outline the configuration details in this subsection. U-Net, U-Net 3+, Res-U-Net, DeepLabV3+, and MA-Net architectures were implemented using the Pytorch framework (Paszke et al. 2019) version 1.9.1, initially loading all networks with weights pre-trained on the Imagenet-1k data set (Russakovsky et al. 2015). The models were trained across 50 epochs with a batch size of 16, 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) (Yi-de et al. 2004) and Lovasz (Berman et al. 2018) loss functions. All experiments were run on a computer with AMD EPYC 7742 64-Core Processor, 1TB RAM, and an A100-SXM4 NVIDIA GPU containing 40GB of memory. The pre-trained weights were fine-tuned on the top layers, aiming to keep the rich features previously learned. In order to improve the variability of the input data, we used the following online training data augmentation techniques: resizing, vertical and horizontal flip, random rotation in intervals of 90 degrees, shift scale rotation, Gaussian noise, Gaussian blur, random brightness contrast, random hue saturation value, optical distortion, grid distortion, and piece-wise affine.

## 4.4 RESULTS AND DISCUSSION

The human data in Fiocruz contains 42 sections from 5 patients. We created the train and test sets in the following manner: 60%-40% train-test split for PAS- and PAMS-stained WSIs, resulting in 9 and 6 WSIs for train and test sets, respectively, for both stains; since we have less HE-stained WSIs, we adopted a 50%-50% train-test split on the HE data set, so that we kept the same test set size across all stains, resulting in 6 WSIs in the train set and 6 in the test one. Considering the mouse data set from Lutnick, we used 78% of WSIs for the training set, leaving the remaining 22% of WSIs for testing. The selection of WSI splits on both data sets was adequately randomized to avoid selection bias.

### 4.4.1 Quantitative analysis

Our first analysis was based on an intra-dataset evaluation, which consisted of training and testing each model on the same data set. The goal was to provide a baseline to com-

pare with the proposed cross-species generalization. Table 4.2 summarizes our findings considering the five selected semantic segmentation networks. Overall, the best results were achieved on Lutnick’s test set, with Dice scores ranging from 0.773 to 0.93. On the other hand, the Dice scores on Fiocruz test set ranged from 0.011 to 0.824. We already expected this difference in generalization because the mice training data is larger and more diverse than the human training set. Still considering the results on Fiocruz test set, it is noteworthy that almost all networks achieved Dice scores above 0.4 (mostly above 0.6) for all stains, except the MA-Net network, which returned scores below 0.02 for all stains. These low scores of MA-Net are justified by the need for a large amount of high-variability data when training attention-based models, which also explains the competitive performance of MA-Net on Lutnick’s test set (0.773). U-Net 3+ returned the highest marks for Lutnick’s test set (0.930) and all stains on Fiocruz test sets (0.772, 0.824, and 0.791, respectively, for HE, PAS, and PAMS). The U-Net 3+ was the only U-Net variation that achieved competitive results for both data sets and for all stains.

From the results in Table 4.2, it is possible to notice that the selected models adequately performed the segmentation of the glomeruli regardless of the numerical differences on each data set. We believe that if we had more data for training, it would guarantee higher scores on the test sets. Now we can move forward to the main focus of our study: the cross-species experiments.

As we ultimately desire to segment human glomeruli, the following evaluation was to compare the results of models trained on human train sets (serving as baseline, directly extracted from Table 4.2, called here as  $H_{train}$ ) with models trained on mouse train sets (shortly here as  $M_{train}$ ), finally testing these models on the Fiocruz test set (shortly here as  $H_{test}$ ). Table 4.3 summarizes the evaluation of the models trained on  $M_{train}$  and  $H_{train}$  sets when predicted (represented by the “ $\rightarrow$ ” symbol) on  $H_{test}$  set. In other words, for each stain, we aim to compare the proposed cross-species generalization ( $M_{train} \rightarrow H_{test}$ ) with the human intra-dataset results ( $H_{train} \rightarrow H_{test}$ ).

For each stain, we highlighted in bold the best results of  $M_{train} \rightarrow H_{test}$  so we can compare them with the best  $H_{train} \rightarrow H_{test}$  score within the respective stain.

Considering a per-stain comparison, HE is the only stain on which the best  $M_{train} \rightarrow H_{test}$  score gets close to the  $H_{train} \rightarrow H_{test}$  result, since Lutnick’s train set contains only HE-stained images. Most HE- $M_{trained}$  networks presented competitive scores, surpassing HE- $H_{trained}$  ones on U-Net, DeepLabV3+, and MA-Net. Considering an overall  $M_{train} \times H_{train}$  comparison (for every stain and network combination), the  $M_{trained}$  models outperformed their respective  $H_{trained}$  ones on 6 out of 15 tests, which is a promising result. However, the high variation between  $M_{trained}$  and  $H_{trained}$  models on PAS and PAMS assessments does not allow us to conclude that our hypothesis holds for different stains. These findings validate our hypothesis on the feasibility of segmenting human glomerulus with a model trained on mouse histology images, at least considering a reasonable amount of mouse data stained on HE. Since the generalization of one stain trained on another one is very different from generalizing one specie to another, one might attempt to explore the DS-FNet network (Silva et al. 2022), which copes with the specific problem of multi-stain generalization by combining segmentation and boundary detection streams.

As we have a limited human data set, we opted to compare the Lutnick’s trained

**Table 4.2** Intra-dataset results (Dice score), considering five segmentation networks over the train and test data sets described on Table 4.1.

Data set	Stain	Network models				
		U-Net	U-Net 3+	Res-U-Net	DeepLabV3+	MA-Net
Lutnick's	HE	0.847	<b>0.930</b>	0.906	0.875	0.773
	HE	0.664	<b>0.772</b>	0.748	0.732	0.018
Fiocruz	PAS	0.457	<b>0.824</b>	0.671	0.416	0.012
	PAMS	0.591	<b>0.791</b>	0.726	0.435	0.011

**Table 4.3** Comparative results (Dice score) considering training the network models on *Lutnick's mouse train set* ( $M_{train}$ ) and on *Fiocruz human train set* ( $H_{train}$ ), separately. The prediction was performed on *Fiocruz human test set* ( $H_{test}$ ). The best results in each  $M_{train} \rightarrow H_{test}$  line is in bold.

Stain	Train $\rightarrow$ Test	Network models				
		U-Net	U-Net 3+	Res-U-Net	DeepLabV3+	MA-Net
HE	$M_{train} \rightarrow H_{test}$	<b>0.747</b>	0.655	0.601	0.738	0.434
	$H_{train} \rightarrow H_{test}^1$	0.664	0.772	0.748	0.732	0.018
	$M_{train} \rightarrow H_{test}$	0.296	0.267	0.273	<b>0.444</b>	0.270
PAS	$H_{train} \rightarrow H_{test}^1$	0.457	0.824	0.671	0.416	0.012
	$M_{train} \rightarrow H_{test}$	0.288	<b>0.401</b>	0.351	0.273	0.159
PAMS	$H_{train} \rightarrow H_{test}^1$	0.591	0.791	0.726	0.435	0.011

<sup>1</sup>Note that this line comes from Table 4.2.

models on both Fiocruz test (whose results directly come from Table 4.3) and the entire (train and test) set, shortly here as  $H_{test}$  and  $H_{entire}$  respectively. Our goal with this analysis is to test our hypothesis in a more influential group, verifying whether the earlier results were biased by the test set random choice. Table 4.4 summarizes the test versus entire set comparison, where we can highlight that the results on  $H_{entire}$  were greater or very close to the  $H_{test}$  results for the most cases, even for the MA-Net network.

In addition, the  $H_{test}$ ' and  $H_{entire}$ ' scores proximity on the HE assessments corroborates our hypothesis validation, strengthening the assumption that human-mice compatibility occurs in intra-stained samples.

**Table 4.4** Comparative results (Dice score) considering training the network models on *Lutnick’s mouse train set*( $M_{train}$ ). Prediction was performed on *Fiocruz human test set*( $H_{test}$ ) and *Fiocruz human train and test sets*( $H_{entire}$ ).

Stain	Test set	Network models				
		U-Net	U-Net 3+	Res-U-Net	DeepLabV3+	MA-Net
	$H_{entire}$	0.708	<b>0.798</b>	0.721	0.735	0.746
HE	$H_{test}^2$	0.747	0.655	0.601	0.738	0.434
	$H_{entire}$	<b>0.554</b>	0.422	0.419	0.392	0.232
PAS	$H_{test}^2$	0.296	0.267	0.273	0.444	0.270
	$H_{entire}$	0.408	0.300	0.383	<b>0.555</b>	0.149
PAMS	$H_{test}^2$	0.288	0.401	0.351	0.273	0.159

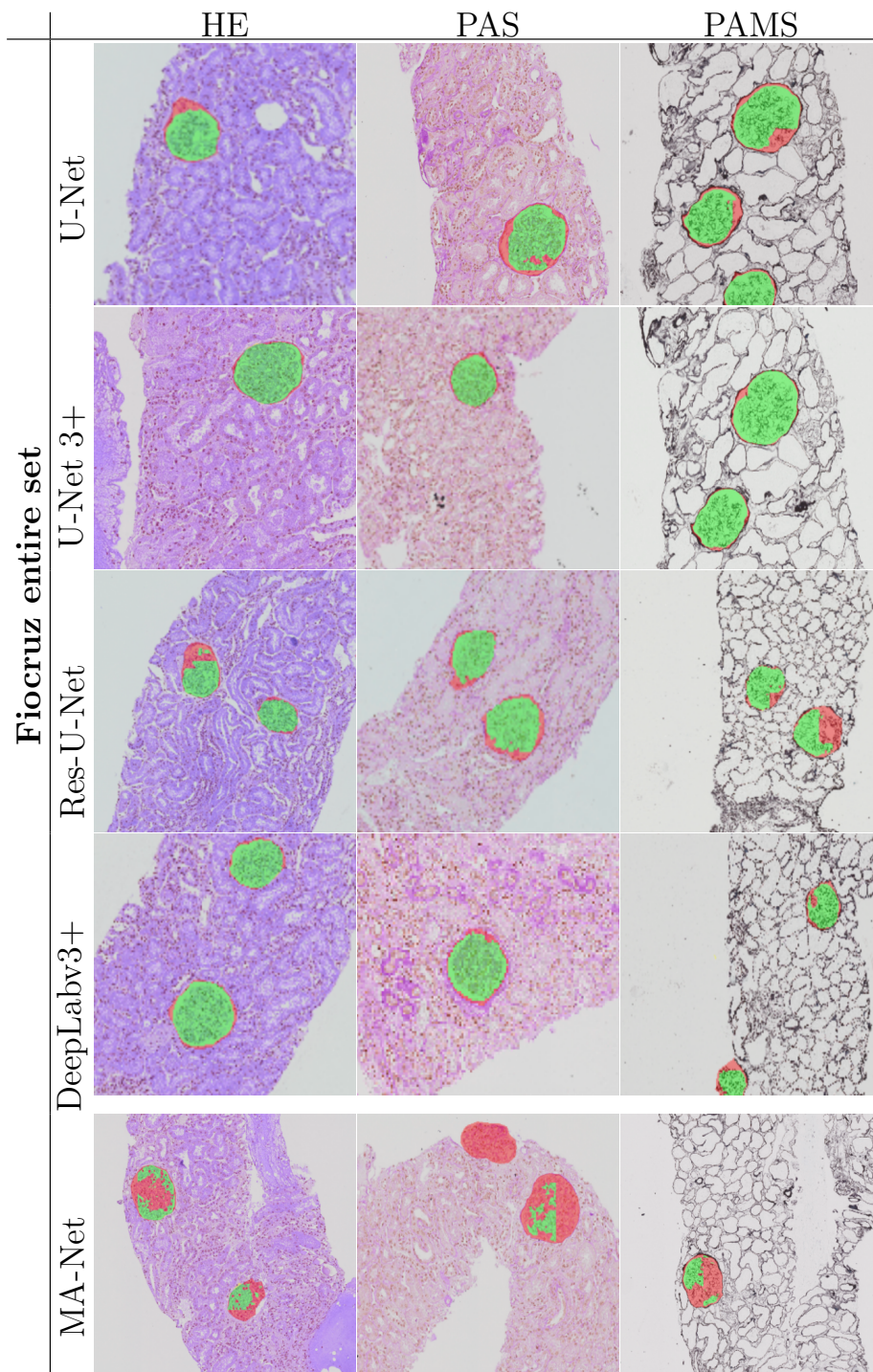
<sup>2</sup>Note that this line comes from Table 4.3.

#### 4.4.2 Qualitative analysis

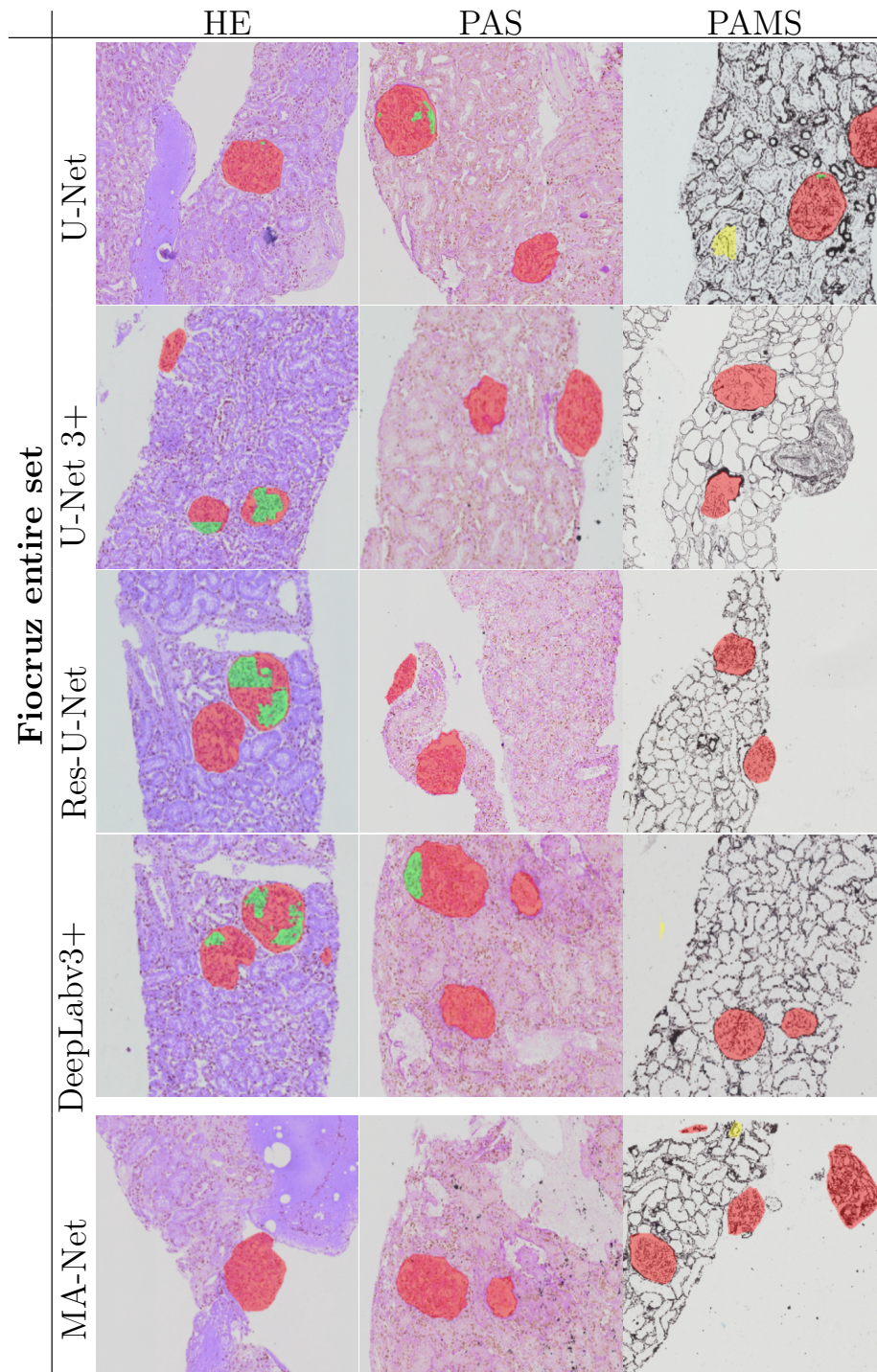
*Experiments over Fiocruz* Figure 4.3 illustrates a visual comparison of the segmentation results over some samples of the Fiocruz data set when trained on Lutnick’s train set. We displayed the true positive predictions in green and false negatives in red over each image in Figure 4.3. Overall, most of the segmenters achieved competitive results, except, as expected, for the MA-Net network, as we have already discussed its underperformance in the quantitative analysis. Most segmenters presented visual results linked to the quantitative analysis on HE-stained images, displaying mostly true-positive segmentation areas (in green). That was an expected outcome, as these networks were also trained over HE-stained images. In addition, we can notice that most of the segmenters also showed a reasonable generalization on PAS- and PAMS-stained images (with slightly better performance for PAS). Considering specific networks, MA-Net presented poor performance for all stains, displaying large false-negative areas. Res-U-Net and U-Net showed better results, but bringing a little more false negatives around glomerular-boundary regions. DeepLabV3+ and U-Net 3+ returned the best predictions, correctly identifying the glomeruli with smaller false-negative areas on glomerular boundaries. We must notice that the U-Net 3+ had a slightly better performance, especially on PAMS-stained images.

In Figure 4.4, some of the visual worst results are depicted. In addition to true positives (green) and false negatives (red), we displayed false positives in yellow over each image. Most results are missing detection (false negatives), but MA-Net and U-Net produced small false positive image regions inside the tissue area on the PAMS-stained images. Additionally, there is a small false positive, where the ‘no glomerulus’ class is incorrectly segmented as glomerulus in an area outside the tissue for DeepLabV3+. In fact, very few false positives can be found in the segmentation of the U-Net 3+, Res-U-Net and DeepLabV3+ over the entire Fiocruz data set.



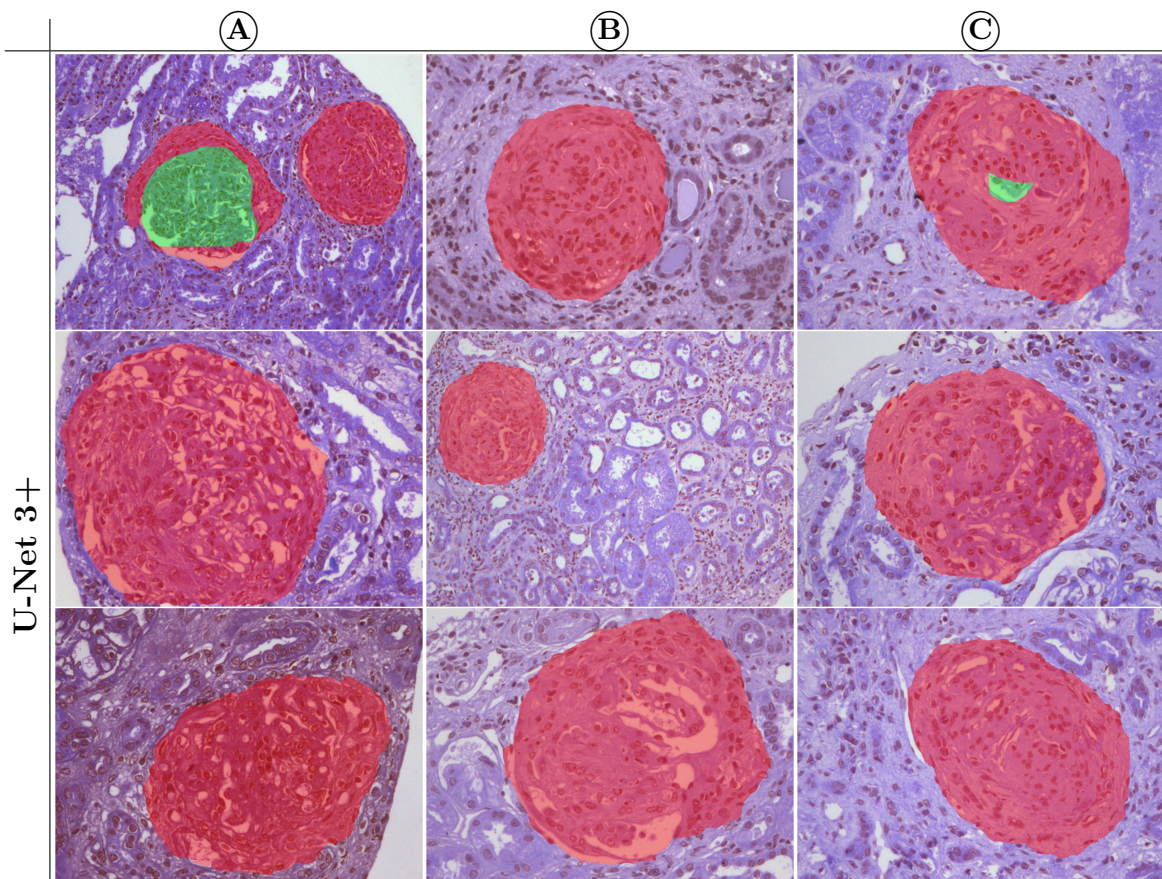


**Figure 4.3** Samples of expected results from U-Net, U-Net 3+, ResUNet, DeepLabV3+, MA-Net, when trained on Lutnick’s train set, predicting on Fiocruz entire data set. Glomeruli in yellow are false positives (very few examples), in red are false negatives, and in green are true positive.



**Figure 4.4** Samples of missing detection from U-Net, U-Net 3+, ResUNet, DeepLabV3+, MA-Net when trained on Lutnick's train set, predicting on Fiocruz entire data set. Note that there is small false-positive regions in yellow on PAMS stain predicted with U-Net, DeepLabv3+, and MA-Net.

*Experiments on unbounded glomeruli.* Within the qualitative analysis, one can also highlight the kind of glomerulus used to assess the performance of the segmenters. Both data sets are indeed comprised of WSIs containing normal glomerulus or those with minimal change disease (MCD), which is a type of lesion that it is not possible to verify a change in the morphological aspect of the glomerulus. Therefore, the anatomy of each glomerulus is preserved, facilitating the segmentation work. The segmentation results would surely be very different in the case of a data set formed, for instance, by glomerulus depicted in Fig. 4.5. All glomeruli in the figure present bruised boundaries due to three different types of lesion: (A) - a cellular crescent filling the Bowman space, (B) - a fibrous crescent and sclerosis of glomerular tufts, and (C) - periglomerular fibrosis. Common to all samples, there is the lack of the Bowman's capsule, which delimits the normal or MCD glomeruli in the data sets we used for performance evaluation. In summary, it is reasonable to state it is not expected that a very well-tuned, off-the-shelf segmentation method would present satisfactory results over the challenging glomeruli illustrated in Fig. 4.5. It is a limitation of our work and an interesting room for investigation in the future.



**Figure 4.5** Samples of segmentation results by U-Net 3+, when trained on Lutnick's train set. This nine images present bruised-boundary glomeruli substantially changed by the following lesions: (A) - a cellular crescent filling the Bowman space, (B) - a fibrous crescent and sclerosis of glomerular tufts, and (C) - periglomerular fibrosis. Note that there is two small true-positive glomeruli region on (A) and (B) lesions.

## 4.5 CLOSURE

This chapter evaluated mouse-to-human transfer learning for glomerulus segmentation. Networks trained on a public HE-stained mouse dataset showed promising results in segmenting human glomeruli (HE, PAS, and PAMS). However, the networks failed to identify glomeruli with indistinct boundaries in the visual analysis. In the next chapter, we discuss the second proposed study for the glomerulus segmentation problem, focusing on glomerulosclerosis lesions.

## SEGMENTING BORDELESS GLOMERULUS IMPAIRED BY GLOMERULOSCLEROSIS

### Contents

---

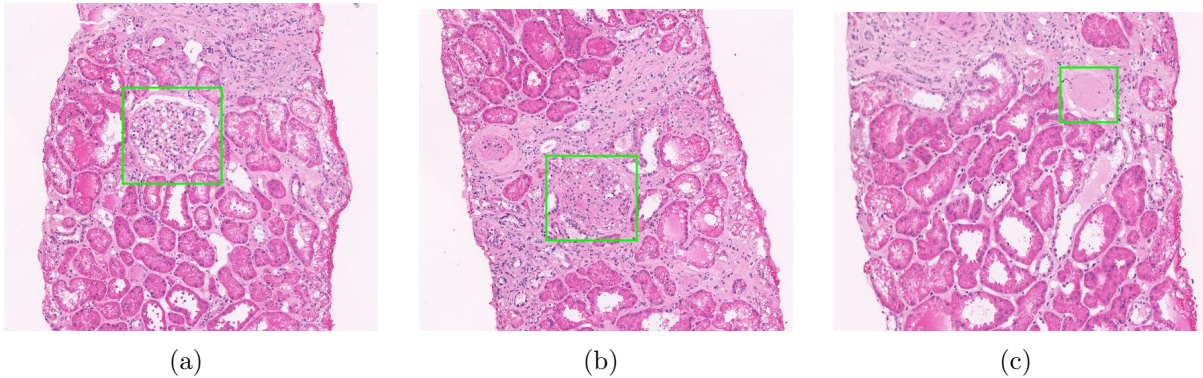
<b>5.1</b>	<b>Background . . . . .</b>	<b>49</b>
<b>5.2</b>	<b>Challenges in segmenting glomerulosclerosis . . . . .</b>	<b>53</b>
<b>5.3</b>	<b>Materials and methods . . . . .</b>	<b>54</b>
5.3.1	Datasets . . . . .	54
5.3.2	Methods . . . . .	54
5.3.3	Implementation details . . . . .	55
<b>5.4</b>	<b>Results and discussion . . . . .</b>	<b>56</b>
5.4.1	Qualitative and quantitative analysis of WSI . . . . .	56
5.4.2	Qualitative and quantitative analysis of cropped glomeruli . . .	59
<b>5.5</b>	<b>Closure . . . . .</b>	<b>61</b>

---

In this chapter, we present our second study, which focuses on segmenting segmental and global glomerulosclerosis, with particular attention to the challenges of segmenting borderless glomeruli in WSIs. We provide a comprehensive overview of the study, detailing the implementation process, experimental analysis, results, and discussion.

### 5.1 BACKGROUND

A kidney biopsy is a medical procedure in which a small specimen of kidney tissue is obtained from a patient for microscopic investigation. This procedure is typically performed to diagnose and follow up on several kidney conditions (e.g., lupus nephritis, IgA nephropathy, and transplant). The histological kidney structure can be analyzed under a microscope or in gigapixel WSIs displayed on computer screens (Niazi et al. 2019; Della Mea et al. 2006, Bayramoglu et al. 2016). Among the structures present in the human



**Figure 5.1** Examples of glomeruli (green bounding box) from the WSI data sets. (a) A healthy glomerulus (normal), (b) A partially sclerotic glomerulus (segmental sclerosis), and (c) A global sclerotic glomerulus.

kidney biopsy, the glomerulus is the nephron unit responsible for blood filtration. Figure 5.1 illustrates examples of glomeruli detected on WSIs: (a) a healthy glomerulus (normal), (b) a partially sclerotic glomerulus (segmental sclerosis), and (c) a global sclerotic glomerulus. Normal and partially sclerotic glomerulus commonly have their Bowman’s space (the white space delineating the glomerulus) intact or partially preserved. In (b), some nuclei (black points) are still seen, making the segmental sclerotic glomerulus keep some special primitives that characterize a healthy glomerulus internally. A globally sclerotic glomerulus is a lesion affecting 100% of the glomerular tuft, marked by fibrosis and prominent scarring or hyaline deposition (Haas et al. 2020). In Figure 5.1 (c), one can find it difficult to distinguish the globally sclerotic glomerulus in the green box from the surrounding renal interstitium. On the other hand, normal glomeruli are characterized by well-preserved internal structures and distinct borders, whereas sclerotic glomeruli exhibit notable abnormalities. Consequently, early attempts to automate glomerulus segmentation have primarily concentrated on identifying normal glomeruli.

Identifying and detecting glomeruli is one of the primary tasks undertaken by pathologists in renal biopsy analysis. Quantitative and qualitative assessments of this task are complex, subjective, and require specialized expertise, often yielding low inter-observer agreement (Haas et al. 2020). The primary objective of automating that task is to provide pathologists with a decision-making tool for pre-screening WSIs (e.g., glomerulus counting), thereby streamlining subsequent automated tasks (e.g., glomerular lesion classification).

Detecting the glomerulus in WSIs is a key step in machine learning-based approaches. Typically, the BC outlines the glomerulus, separating it from surrounding tissue. However, in cases of global sclerosis, the glomerulus often lacks a visible BC, making it appear borderless.

Many studies address the problem of glomerulus segmentation (Bel et al. 2018; Marsh et al. 2018; Hermsen et al. 2019; Bueno et al. 2020; Altini et al. 2020; Jiang et al. 2021; Davis et al. 2021; Gallego et al. 2021). Most of these studies do

not differentiate between types of glomeruli (Gadermayr et al. 2019; Jha et al. 2021; Jayapandian et al. 2021; Bouteldja et al. 2021; Silva et al. 2022; Souza et al. 2023; Altini et al. 2023) , and the performance of the proposed machine learning methods used for segmentation is evaluated by disregarding the various lesions that can affect the glomeruli. Consequently, the performance of the proposed methods may be overestimated, especially when considering their translation to routine clinical practice. To further understand the current landscape of glomerulus segmentation research, Table 5.1 summarizes the works related to our study. The criteria for selecting these works were threefold: (i) perform glomerulus segmentation, (ii) utilize WSIs, and (iii) be available in PubMed (<https://pubmed.ncbi.nlm.nih.gov>).

Several insights can be drawn from Table 5.1: U-Net and its variations are a prominent choice for glomerulus segmentation; the most commonly used stains are HE, PAS, and PAMS, with only three studies using TRI, and our study is one out of two to consider all three glomerulus classes: (i) normal, (ii) segmental sclerosis, and (iii) global sclerosis. It is also noteworthy that most of the datasets used are private. Due to the limited availability of public datasets, the time-consuming annotation process by pathologists, and the low categorization of sclerosed glomeruli, most studies develop datasets with fewer than 100 WSIs and fewer than 3,000 glomeruli, usually including only two glomerulus classes: normal and sclerosed. This distinction in approach is further highlighted by the fact that many studies do not differentiate between types of glomeruli, aiming instead to develop a robust segmenter for a canonical structure (Gadermayr et al. 2019; Jha et al. 2021; Jayapandian et al. 2021; Bouteldja et al. 2021; Altini et al. 2023; Silva et al. 2022; Souza et al. 2023). Except (Silva et al. 2022), which proposes its own DL network applied to a more complex segmentation problem, the remaining studies employ off-the-shelf (or eventually modified) methods, primarily on WSIs stained with HE, PAS, and PAMS and considering mostly normal glomeruli.

Some related works consider two classes of glomeruli when tackling the problem of segmenting sclerotic glomeruli: normal and with sclerosis (Bel et al. 2018; Marsh et al. 2018; Hermsen et al. 2019; Bueno et al. 2020; Altini et al. 2020; Davis et al. 2021; Gallego et al. 2021). These studies often do not specify the performance of the segmenters for each glomerulus category, leaving unclear which Dice or IoU scores were achieved for normal versus sclerotic glomeruli. While the reported results are generally high, they are likely overestimated due to the common issue of unbalanced datasets, which typically contain many more normal glomeruli than sclerotic ones.

Only the study by Jiang et al. (2021) investigates glomeruli with global sclerosis alongside two other classes: normal glomeruli and glomeruli with other lesions (moderate to severe mesangial hypercellularity or expansion, crescents, and apparent endothelial proliferation). The methodology used by Jiang et al. (2021) involves cropping the WSIs into 2048 x 2048 pixel patches, which are manually verified to ensure that all irrelevant areas are filtered out and that each glomerulus remains entirely within a patch. This meticulous approach clearly impacts the segmenter’s performance, as demonstrated in the present study. For experimental results, the authors used the training and test datasets within the same glomerulus class domain, with the mIoU metric employed to evaluate

**Table 5.1** Summary of related work for glomerulus segmentation. Dashes indicate instances where information was unreported.

Year	Reference	#WSI	#Glom.	Method	Stain	Classes	Data set
2018	(Bel et al. 2018)	15	944	FCN M-FCN U-Net	PAS	Normal, Sclerosis	Private
2018	(Marsh et al. 2018)	48	3,867	CNN(VGG-16) LoG Blob-detection Linear Regression	HE	Normal, Sclerosis	Private
2019	(Gadermayr et al. 2019)	24	–	SW-CNN U-Net-S U-Net-D	PAS	–	Private
2019	(Hermsen et al. 2019)	60	238	U-Net	PAS	Normal, Sclerosis	Private
2020	(Bueno et al. 2020)	47	1,245	U-Net (SegNet + AlexNet)	PAS	Normal, Sclerosis	Private
2020	(Altini et al. 2020)	26	2,772	SegNet DeepLabv3+	PAS	Normal, Sclerosis	Private
2021	(Jha et al. 2021)	61	1,334	Mask-RCNN U-Net DeepLabv3	HE PAS PAMS	–	Private
2021	(Jiang et al. 2021)	348	8,665	Cascade Mask R-CNN	PAS PAMS	Normal, Global Sclerosis, Other lesions	Private
2021	(Jayapandian et al. 2021)	459	1,196	U-Net	HE PAS PAMS TRI	–	Private
2021	(Bouteldja et al. 2021)	168	2,611	Modified U-Net	PAS	–	Private
2021	(Davis et al. 2021)	258	24,133	U-Net	HE	Normal, Sclerosis	Public
2021	(Gallego et al. 2021)	51	2,429	U-Net	HE PAS	Normal, Sclerosis	Public
2022	(Silva et al. 2022)	665	5,309	DS-FNet	HE PAS PAMS TRI	–	Public Private
2023	(Souza et al. 2023)	60	1,430	U-Net U-Net3+ Res-U-Net DeepLabV3+ MA-Net	HE PAS PAMS	–	Public Private
2023	(Altini et al. 2023)	875	23,477	Cascade R-CNN	PAS	–	Private
<b>2024</b>	<b>Ours</b>	<b>83</b>	<b>800</b>	<b>U-Net</b> <b>U-Net3+</b>  <b>SwinTransf.+U-Net</b>  <b>Fine-tuned versions</b> <b>SegGPT</b>	<b>HE</b> <b>PAS</b>  <b>PAMS</b>	<b>Normal,</b> <b>Segmental</b> <b>sclerosis,</b> <b>Global scler-</b> <b>osis</b>	<b>Private</b>

the three classes. The best scores were achieved for the PAS-stained normal class with a mIoU of 0.943, the PAMS-stained glomeruli with other lesions class with 0.812, and the PAMS-stained and PAS-stained global sclerosis class with 0.692. When evaluating the segmentation performance of WSIs using the mIoU metric, the results were lower than those obtained by patch-based evaluation. The best scores were found for the PAMS-



stained normal class with 0.742, the PAMS-stained glomeruli with other lesions class with 0.660, and the PAS-stained global sclerosis class with 0.646.

## 5.2 CHALLENGES IN SEGMENTING GLOMERULOSCLEROSIS

It is important to consider that the work by Jiang et al. (2021) does not represent the most likely situation to be found in the real-world scenario of renal pathology diagnosis since it deals with glomerulus manually cropped from WSIs.

It is important to note that the study by Jiang et al. (2021) does not reflect the most realistic scenario in renal pathology diagnosis, as it involves manually cropping glomeruli from WSIs. Instead, the ideal approach would involve segmenting the glomerulus directly on the WSI. Segmenting the entire glomerulus in cropped images is simpler because the glomerulus is more prominent compared to other renal structures (such as the interstitium, tubules, and vessels) and is often zoomed in for a focused view. An approach that emphasizes the analysis of WSI from the start will provide more efficient support for pathologists' decision-making. Unlikely, we have investigated glomerulus segmentation starting from WSIs and subsequently assessed automatically cropped glomeruli. Both WSIs and cropped were evaluated using the mean-Dice. It is important to emphasize that in our study, the non-foundation models were trained with a limited number of samples exhibiting segmental and global glomerulosclerosis, particularly during fine-tuning. This approach was intended to mirror the real-world challenges of annotating sclerotic glomeruli, which are compounded by the lack of consensus among pathologists when defining masks for glomeruli with severely deteriorated Bowman's capsules.

Whereas early works have investigated the task of glomerulus segmentation by initially focusing on normal glomeruli (using a canonical structure as the starting point), addressing the issue of global sclerosis remains a recurrent challenge in the pathologists' routine. The main contribution of this chapter is then to elucidate this ill-posed problem, 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 carried out a set of experiments conducted on a dataset of histological images of representative cases of the diagnostic routine, containing a variety of glomeruli, either normal or with sclerotic lesions.

The primary challenge with global glomerulosclerosis is accurately delineating their boundaries in histopathological images. This difficulty arises because they may exhibit similarities with another renal structure, the interstitium, often sharing similar color tones between these two structures (see Fig. 5.1 (c)). In the next section, we begin addressing the problem by assessing how robust it is to segment normal glomeruli and further analyzing the critical aspects of segmenting borderless glomeruli affected by global sclerosis lesions.

### 5.3 MATERIALS AND METHODS

The dataset compilation used in this paper originates from two newly introduced sets named Normal and Sclerosis. The “Normal” dataset extends the prior dataset found in (Souza et al. 2023), now featuring 45 WSIs of human kidney biopsies. The “Sclerosis” dataset is newly created and includes 37 human kidney sections. The “Mixed” dataset is also new and contains 1 human kidney section. 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 from a VS 110 Olympus scanner with 40× magnification. The main characteristics of these two datasets are summarized in Table 5.2.

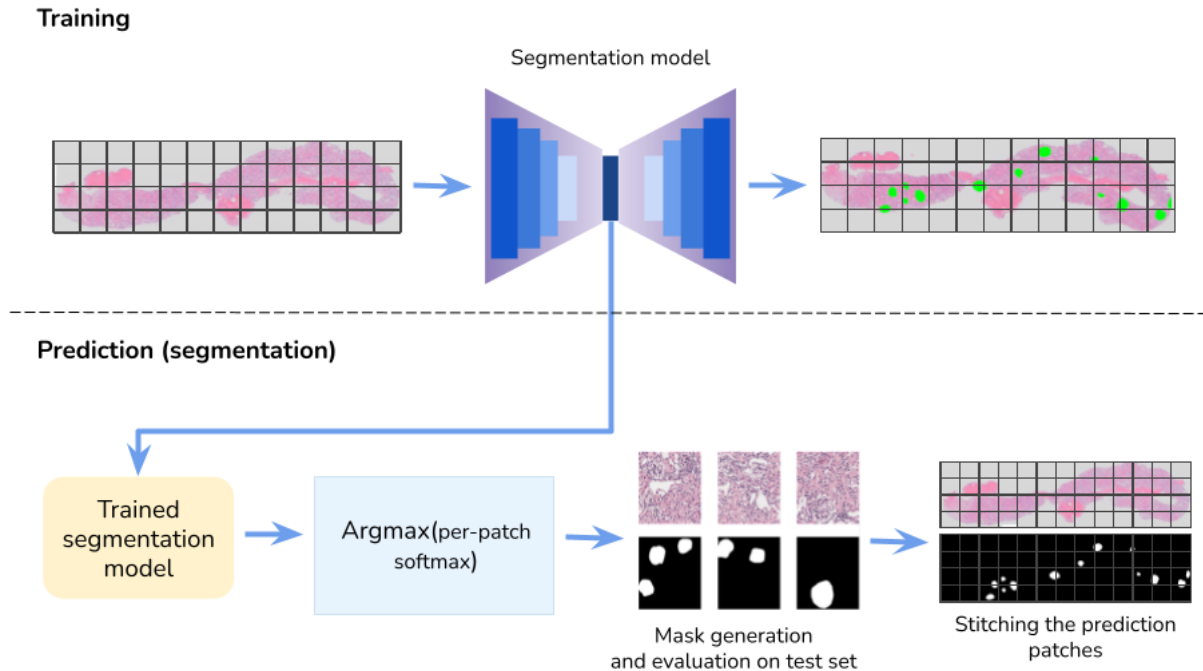
#### 5.3.1 Datasets

#### 5.3.2 Methods

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). The proposed framework to address this challenge is illustrated in Fig. 5.2 and comprises training and prediction phases. Since this framework is agnostic regarding the segmentation network, based on the literature review, we employed four powerful methods: vanilla U-Net (Ronneberger et al. 2015), U-Net 3+ (Huang et al. 2020), Swin-Transformer (Liu et al. 2021) as the backbone of a U-Net, and Segmentation Generative Pre-trained Transformer (SegGPT) (Wang et al. 2023). SegGPT is a generalist model developed for context-aware segmentation tasks, utilizing a foundation model that distinguishes it from the other non-foundation segmentation methods. It approaches training as an in-context coloring task, applying random color mapping to each data sample. We adapted this model for gigapixel histopathological images of glomeruli, allowing it to function without a conventional training phase and thus avoiding the need for extensive task-specific training or fine-tuning. Instead, SegGPT leverages feature representations from a few query samples during inference. There is no need to manually crop the input WSI into patches; each segmenter predicts object masks directly on the whole-slide image. A stitching algorithm is then employed to seamlessly reassemble the patches into the final WSI output. To evaluate this approach, our team of pathologists assembled a

**Table 5.2** Summary of the characteristics of the datasets.

Dataset	Stain	#WSI		#Glom.		#Glom. by class			Average Pixels (height × width)
		Train	Test	Train	Test	Normal	Segmental	Global	
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
Mixed	HE	1	-	12	-	4	4	4	26,624 × 13,824



**Figure 5.2** 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.

specialized dataset of WSIs, with further details provided in the following section.

### 5.3.3 Implementation details

Based on the articles presented in Table 5.1 (see Section 5.1), the initial experiments in our work, and glomeruli segmentation in the Kaggle competition (Howard et al. 2020), we provide a detailed description of the implementation parameters in this subsection. U-Net, U-Net 3+, SwinTransformer + U-Net, and SegGPT methods were implemented using the PyTorch framework (Paszke et al. 2019) version 1.9.1. The initial setup involved loading all networks with weights pre-trained on the Human BioMolecular Atlas Program (HuBMAP) dataset (Howard et al. 2020). 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 BCE (Yi-de et al. 2004) and Lovasz (Berman et al. 2018) loss functions. The WSIs were automatically tiled into patches of 1,024 x 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 x 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 improve the variability of the input data, we used the following online training data augmentation techniques: resizing, vertical and horizontal flip, random rotation in intervals of 90 degrees, shift scale rotation, Gaussian noise, Gaussian blur, random brightness contrast, and random hue saturation value.

## 5.4 RESULTS AND DISCUSSION

WSIs analyses significantly shape the daily routine of a pathologist. For an expert system to assist these professionals effectively in day-to-day diagnosis practice, it must handle gigapixel images containing numerous renal structures and various lesions, which often obscure the canonical appearance of these structures. Hence, the primary goal of this study is to work within the framework depicted in Fig. 5.2. Given that normal glomeruli 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. 5.1 (b)). However, this convenience does not necessarily extend to glomeruli affected by global sclerosis (see Fig. 5.1 (c)), even when a fine-tuning strategy is applied to samples of global sclerotic glomeruli. The secondary objective is to demonstrate that segmenting a crop containing an entire globally sclerotic glomerulus can lead to overestimation. This evaluation was conducted similarly to the approach proposed by (Jiang et al. 2021).

### 5.4.1 Qualitative and quantitative analysis of WSI

Tables 5.3 and 5.4 summarize the results found by our proposed study, considering the splits defined in Table 5.2. The metric used to assess the performance was mDice. Table 5.3 shows the evaluation of the behavior of the three segmenters by training on normal glomeruli and subsequently testing on either normal or segmental sclerotic glomeruli. The results show that any of the network models were able to reach state-of-the-art performance (see Table 5.3), 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 segmental sclerotic glomerulus (considering U-Net and SwinTransformer + U-Net; average mDice of 0.730 and 0.702, respectively). However, the performance of the three trained deep learning models, U-Net, U-Net3+, and SwinTransformer + U-Net, was significantly poor in segmenting globally sclerotic glomeruli, regardless of whether they were fine-tuned or not in the WSI segmentation task. As shown in Table 4, both U-Net3+ and SwinTransformer + U-Net recorded a mDice score of zero when trained on normal glomeruli and subsequently tested on globally sclerotic glomeruli. In contrast, significant performance improvements were achieved with the foundation model for globally sclerotic glomeruli, with SegGPT reaching an average mDice score of 0.428.

These results make clear that the primary factor for achieving state-of-the-art performance is dataset unbalance. This is due to the small number of glomeruli with global sclerosis found in the WSI (see Table 5.2). The numbers achieved by testing on WSI regarding only global sclerosis are invariably tiny, whether fine-tuning the models with

**Table 5.3** Comparison of Dice scores and their standard deviations obtained from three segmentation networks trained on the Normal’s train set. Predictions were performed on both the Normal’s test set and the Sclerosis’ test set as described in Table 5.2

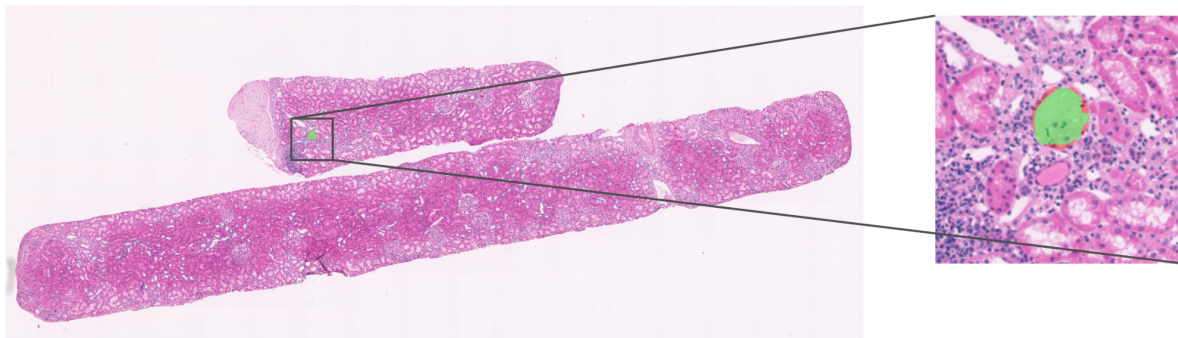
Input	Class	Stain	mDice (test subset)		
			U-Net	U-Net3+	SwinTransformer+U-Net
WSI	Normal	HE	0.923( $\pm 0.014$ )	0.864( $\pm 0.022$ )	<b>0.935(<math>\pm 0.008</math>)</b>
		PAS	0.915( $\pm 0.005$ )	<b>0.931(<math>\pm 0.022</math>)</b>	0.907( $\pm 0.009$ )
		PAMS	<b>0.937(<math>\pm 0.006</math>)</b>	0.926( $\pm 0.006$ )	0.850( $\pm 0.048$ )
	Segmental	ALL	<b>0.919(<math>\pm 0.018</math>)</b>	0.908( $\pm 0.028$ )	0.918( $\pm 0.033$ )
		HE	<b>0.720(<math>\pm 0.192</math>)</b>	0.575( $\pm 0.131$ )	0.650( $\pm 0.150$ )
		PAS	<b>0.730(<math>\pm 0.197</math>)</b>	0.643( $\pm 0.184$ )	0.702( $\pm 0.156$ )
		ALL	<b>0.728(<math>\pm 0.163</math>)</b>	0.667( $\pm 0.163$ )	0.675( $\pm 0.186$ )

**Table 5.4** Comparison of Dice scores and their standard deviations obtained from four segmentation networks tested exclusively on global sclerotic glomeruli using both the input WSI and per-crop data from the Sclerosis’ test set. Columns 3-5 list the results for three segmenters without fine-tuning, trained on the Normal’s train set. Columns 6-8 show the results for the same three segmenters with fine-tuning using the pre-trained model from the Normal train’s set and additional training from the Mixed’s dataset. Column 9 lists the results for the SegGPT segmenter, which uses feature representations (without a training set) by querying the Mixed’s dataset.

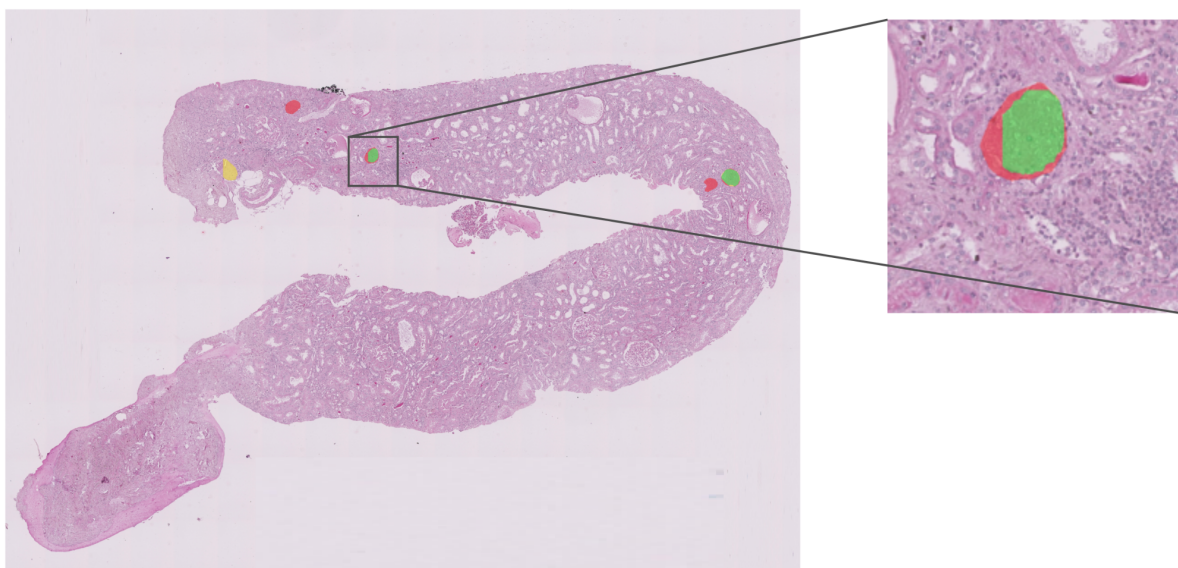
Input	Stain	mDice (test subset)						
					Fine-tuning			SegGPT
		U-Net	U-Net3+	SwinTransf. +U-Net	U-Net	U-Net3+	SwinTransf. +U-Net	
WSI	HE	0.020 ( $\pm 0.018$ )	0.000 ( $\pm 0.000$ )	0.000 ( $\pm 0.000$ )	0.066 ( $\pm 0.054$ )	0.000 ( $\pm 0.000$ )	0.000 ( $\pm 0.000$ )	<b>0.428</b> ( $\pm 0.258$ )
	PAS	0.002 ( $\pm 0.001$ )	0.027 ( $\pm 0.022$ )	0.017 ( $\pm 0.011$ )	0.011 ( $\pm 0.009$ )	0.151 ( $\pm 0.039$ )	0.131 ( $\pm 0.041$ )	<b>0.366</b> ( $\pm 0.261$ )
	ALL	0.005 ( $\pm 0.002$ )	0.020 ( $\pm 0.013$ )	0.004 ( $\pm 0.002$ )	0.041 ( $\pm 0.013$ )	0.124 ( $\pm 0.033$ )	0.034 ( $\pm 0.015$ )	<b>0.400</b> ( $\pm 0.255$ )
	Patch	HE	0.000 ( $\pm 0.000$ )	0.000 ( $\pm 0.000$ )	0.035 ( $\pm 0.164$ )	0.420 ( $\pm 0.142$ )	0.386 ( $\pm 0.108$ )	0.112 ( $\pm 0.110$ )
	PAS	0.087 ( $\pm 0.268$ )	0.225 ( $\pm 0.282$ )	0.002 ( $\pm 0.010$ )	0.448 ( $\pm 0.201$ )	0.461 ( $\pm 0.126$ )	0.224 ( $\pm 0.203$ )	<b>0.654</b> ( $\pm 0.363$ )
	ALL	0.084 ( $\pm 0.265$ )	0.059 ( $\pm 0.204$ )	0.001 ( $\pm 0.009$ )	0.587 ( $\pm 0.284$ )	0.330 ( $\pm 0.150$ )	0.121 ( $\pm 0.114$ )	<b>0.701</b> ( $\pm 0.354$ )

HE-stained patches or not. It is noted that when confronted with a real-world dataset, the networks cannot even locate the glomeruli with global sclerosis, as indicated by a Dice virtually equal to zero in most cases (see Table 5.4). Jiang et al. (2021) cope with this problem by manually cropping individual glomerulus from the WSI and presenting the patches to the network.

Conversely, promising results were obtained using SegGPT, as demonstrated in Table 5.4, with visual outputs on WSIs shown in Figure 5.3. Unlike non-foundation models, SegGPT is not trained in the traditional manner but instead relies on receiving a few image-mask pairs to query new image classes within the WSI. For our evaluation, we utilized glomeruli from the Mixed dataset to perform the query and subsequently tested SegGPT on the global sclerosis subset. In the HE-stained image (top), the model performed well, correctly segmenting a glomerulus (highlighted in green). In the PAS-stained image (bottom), the segmentation was reasonable, with one true positive glomerulus (in



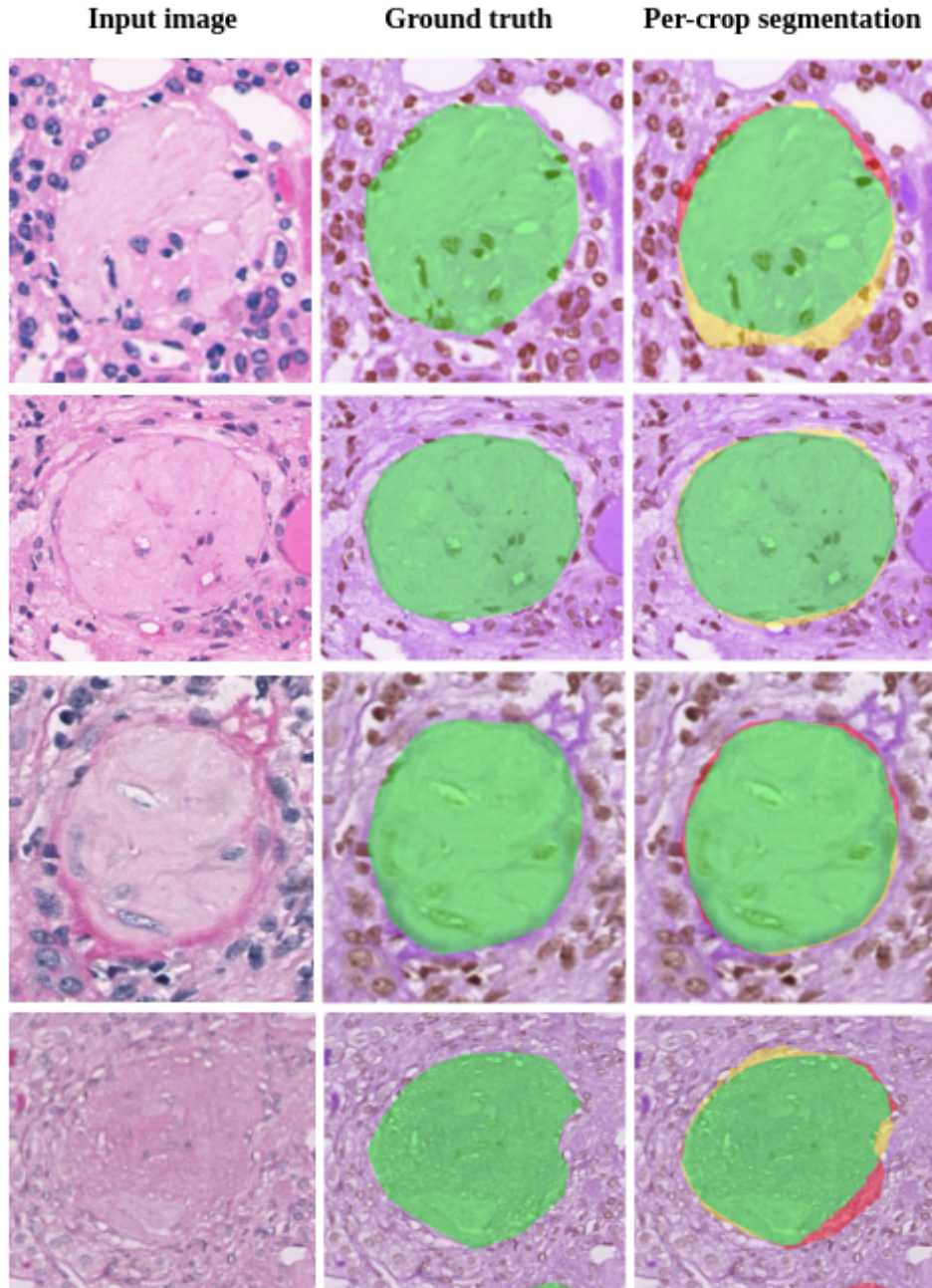
(a)



(b)

**Figure 5.3** Some visual results from SegGPT’s WSI segmentation for global sclerosis. (a) HE-stained WSI and a zoomed patch: The WSI segmentation shows the only globally sclerotic glomerulus (green mask) with true-positive pixels. The zoomed patch highlights detailed true-positive pixels (in green) and very few false negatives (in red). (c) PAS-stained WSI and a zoomed patch: The WSI displays a yellow mask representing false-positive pixels, two red masks indicating false-negative pixels, and two green masks highlighting true-positive pixels. The zoomed patch shows two glomeruli: one with missed detection (red mask) and the other containing true-positive (green mask), false-positive, and false-negative pixels, with true-positive pixels predominating.

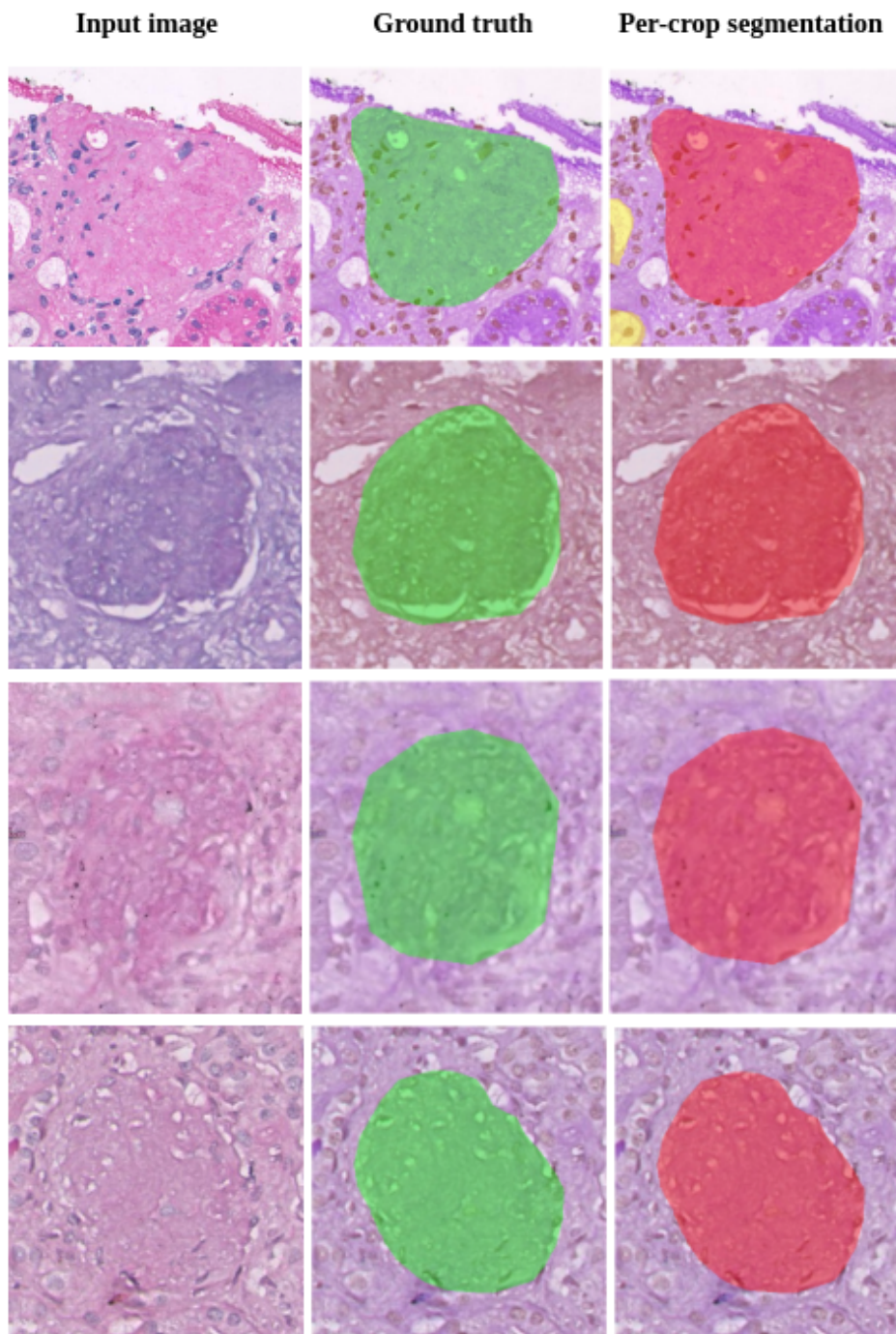
green), one false positive (in yellow), and two missed detections (in red).



**Figure 5.4** Samples of expected results from SegGPT applied on Global Sclerosis’ test set. From left to right: input image, ground truth, and results from SegGPT’s per-crop segmentation for global sclerosis. Yellow masks represent false-positive pixels, red masks indicate false-negative pixels, and green masks show true-positive pixels (best viewed in color). The first two rows are HE-stained images, while the last two rows are PAS-stained images.

#### 5.4.2 Qualitative and quantitative analysis of cropped glomeruli

In Table 5.4, results were obtained by extending our framework to include global sclerotic glomeruli from the per-crop input. Our best performance was achieved with the



**Figure 5.5** Samples of missing detection from SegGPT predicting on Global Sclerosis' test set. From left to right: input image, ground truth, and results from SegGPT's per-crop segmentation for global sclerosis. Yellow masks represent false-positive pixels, red masks indicate false-negative pixels. The first two rows are HE-stained images, while the last two rows are PAS-stained images



SegGPT foundation model, yielding a Dice score of 0.742. The U-Net, U-Net3+, and SwinTransformer + U-Net models achieve average Dice scores of 0.587, 0.461, and 0.224, respectively. In these experiments, we fine-tuned the non-foundation models using the pre-trained model generated by the initial framework along with the Mixed dataset of per-crop glomerulus, which was used exclusively for training (see Datasets). The Sclerosis test set consisted exclusively of globally sclerotic glomeruli. It is important to note that when the trained non-foundation models were applied to these globally sclerotic glomeruli without fine-tuning, the results were zero or close to zero across U-Net, U-Net3+, and SwinTransformer + U-Net.

According to the quantitative results listed in Table 5.4, the best segmenter was SegGPT applied to per-crop data, which generated the qualitative results illustrated in Fig. 5.4. All the stains show similar visual results on the samples depicted in the figure, with prominent true positive segmentations (in green), minor false positives (in yellow), and very few false negatives (in red). In contrast, Figure 5.5 highlights some of the worst visual results obtained using the SegGPT segmenter. In addition to false negatives (in red), false positives (in yellow) are highlighted in the figure, with no true positives detected. All visual results from the per-crop glomerulus were characterized by missed detections (false negatives). Additionally, a few false positive regions were generated outside the glomerulus area in the HE-stained image (initial per-crop segmentation image).

## 5.5 CLOSURE

This chapter evaluates the challenges associated with segmenting globally sclerotic glomeruli in WSIs. Although the models demonstrate strong performance for normal and partially sclerotic glomeruli, their efficacy significantly declines when addressing globally sclerotic cases, even with fine-tuning. However, the SegGPT foundation model shows promise, providing substantial improvements in segmentation accuracy. In the next chapter, we discuss the results shown throughout this thesis. We conclude this thesis by presenting our final considerations and perspectives for future work in this research line.



## DISCUSSION AND CONCLUSIONS

### Contents

---

<b>6.1 Overall results in the experiments . . . . .</b>	<b>63</b>
6.1.1 Mouse-to-human transfer learning . . . . .	63
6.1.2 The problem of segmenting global glomerulosclerosis . . . . .	64
<b>6.2 Strengths and limitations . . . . .</b>	<b>64</b>
<b>6.3 Contribution and future work . . . . .</b>	<b>65</b>

---

This thesis explores the problem of glomerulus segmentation in high-resolution histopathological images, trained on limited data. To address that, we developed two studies: cross-species analysis between mice and humans, and segmentation of sclerotic glomeruli using a foundation model. These studies are based on transfer learning and few-shot learning, respectively. In this chapter, we present key insights from these studies by analyzing the overall experimental results, highlighting the strengths and limitations of the investigated methods, and outlining potential directions for future research in this field.

### 6.1 OVERALL RESULTS IN THE EXPERIMENTS

#### 6.1.1 Mouse-to-human transfer learning

We proposed investigating the feasibility of segmenting glomeruli in human WSIs by training deep-learning models on mouse data only. A set of semantic segmenters was evaluated, with their performance assessed on two datasets consisting of 18 mouse WSIs and 42 human WSIs.

The experiments demonstrated that our hypothesis is indeed feasible, as most networks, when trained on a public HE-stained mice data set, achieved promising results in the human glomerulus segmentation task (including samples stained with HE, PAS, and PAMS). However, we found that generalization was only achievable when the training

and test sets contained images with the same staining. In the cross-species scenario, testing on PAS and PAMS-stained images resulted in low and inconsistent scores.

It is important to note that these conclusions are based on general-purpose segmentation networks, like those used in our experiments. Additionally, both the human and mouse datasets contained well-defined glomeruli (with no morphological lesions), which facilitated direct generalization. To test this shape assumption, we collected nine glomerular images with irregular boundaries and observed that the top-performing network (U-Net 3+) was almost unable to identify any glomeruli in the visual analysis. Although all networks successfully segmented intra-stain images, we could not confirm the same performance on inter-stain ones.

### 6.1.2 The problem of segmenting global glomerulosclerosis

We investigated glomerulus segmentation in WSIs and cropped samples of human kidney tissue, focusing on the challenges of segmenting sclerotic glomeruli. A set of semantic segmenters was evaluated, with their performance assessed on three datasets: 45 WSIs of normal glomeruli, 37 WSIs of glomerulosclerosis, and 1 WSI containing both normal and sclerotic glomeruli.

Our findings indicated that while segmenting glomeruli with well-preserved structures is largely a solved problem, the deterioration of Bowman’s capsule presents significant challenges for existing segmentation methods. Specifically, conventional supervised models (non-foundation) struggled to segment globally sclerotic glomeruli, as the lack of distinct boundaries resulted in near-zero performance.

In contrast, we identified a promising alternative in the SegGPT foundation model. Our results demonstrated that SegGPT achieved over 40% segmentation accuracy for globally sclerotic glomeruli in WSIs — a significant improvement over the near-zero performance of non-foundation models, especially given that SegGPT requires neither extensive training nor fine-tuning.

## 6.2 STRENGTHS AND LIMITATIONS

Non-foundation models (classical deep learning) and foundation models are essential methods for addressing the proposed task.

For non-foundation models, we can highlight two positive points:

- (i) Satisfactory performance in the initial study on mouse-to-human transfer learning using five methods: U-Net, U-Net3+, Res-U-Net, DeepLabv3+, and MA-Net. The first four achieved results above 60%, while the last method obtained approximately 43%;
- (ii) Acceptable performance in the second study addressing the issue of borderless glomeruli with segmental glomerulosclerosis, achieving over 72%. However, the results for global glomerulosclerosis remained notably poor, even when employing a fine-tuning strategy.

Regarding foundation models, several key points emerge:

- (i) The SegGPT foundation model shows promise, delivering significant improvements in segmenting glomeruli completely borderless;
- (ii) SegGPT achieved considerable enhancement in segmenting globally sclerotic glomeruli by utilizing only a few query samples, with a significantly higher mDice(> 0.37).

Non-foundation models and foundation models were applied to input WSIs and cropped samples to investigate global glomerulosclerosis lesions. When we used non-foundation models with fine-tuning, we observed minimal improvements in the results. This led us to formulate our hypothesis of evaluating a foundation model utilizing only a few query samples to investigate whether borderless glomeruli could be labeled.

In our two comprehensive studies, we identified some limitations:

- (i) In the first study, cross-species generalization was only achievable when both training and testing sets contained images with the same staining type. When testing on PAS and PAMS stained samples, we observed low and variable scores in the cross-species scenario. Additionally, there were limitations in segmenting borderless glomeruli. In supplementary experiments, we collected nine glomerular images with damaged contours representing three types of lesions: (i) Cellular crescents filling Bowman’s space, (ii) Fibrous crescents with sclerosis of glomerular tufts, and (iii) Periglomerular fibrosis. We found that the network with the highest score (U-Net 3+) almost failed to detect glomeruli upon visual analysis.
- (ii) In the second study, using non-foundation models, we achieved only satisfactory results with models trained exclusively on normal glomeruli (using HE, PAS, and PAMS-stained images) demonstrated strong performance for normal glomeruli (mDice > 0.92) and moderate performance for partially sclerotic glomeruli (mDice > 0.72). However, their performance dropped drastically for globally sclerotic glomeruli (mDice > 0.02), with negligible improvements even after fine-tuning on globally sclerotic samples. In contrast, SegGPT achieved substantial improvement on globally sclerotic glomeruli by leveraging only a few query samples, with a significantly higher mDice(> 0.37).

### 6.3 CONTRIBUTION AND FUTURE WORK

In light of the empirical evidence provided throughout this thesis, we are now able to answer the questions raised in Chapter 1. They are:

- (i) Is it possible to design an end-to-end CNN for semantic segmentation of glomeruli in high-resolution images?

Yes, it is. In this work, we developed a methodology that spans from data acquisition to stitching the predicted patches in our two studies presented in Chapters 4 and 5. The experiments demonstrate satisfactory results for both normal glomeruli and two types of sclerotic lesions: segmental and global.

- (ii) Is it possible to segment human glomerulus from WSIs by using a model trained with only mouse glomerular images?

Yes, it is possible to achieve successful segmentation of glomeruli trained on mouse data, based on the morphological similarities between mouse and human glomerulus. We evaluated several state-of-the-art segmentation networks for our proposed cross-species generalization discussed in Chapter 4. Using mice glomeruli images offers significant advantages, including the relative ease of obtaining samples compared to human samples. In a laboratory setting, mice can be bred and monitored to ensure they are either healthy or have specific lesions, and their kidneys can be harvested for targeted research. Additionally, the smaller size of mouse kidneys, typically around 1.25 to 1.5 cm in length, is another advantage (Wang et al. 2017). In contrast, human kidney samples are generally limited to small biopsy sections, as human kidneys are larger, measuring about 8 to 14.5 cm in length (Rao et al. 2013).

- (iii) How do we segment borderless glomeruli?

Using traditional deep learning methods, we achieved satisfactory results with partially borderless glomeruli (segmental sclerosis). However, segmenting fully borderless glomeruli (global sclerosis) presented significant challenges, as the lack of clear boundaries and limited training data resulted in near-zero performance. The SegGPT foundation model provides an alternative approach for identifying and segmenting globally sclerotic glomeruli in both WSI and per-crop evaluations, where we obtained mDice scores exceeding 0.37 and 0.65, respectively. Unlike traditional models, the SegGPT model requires only a few sample images for guidance, demonstrating potential for enhanced segmentation accuracy. In Chapter 5, we discuss the challenges of segmenting these borderless glomeruli in WSIs, underscoring the promise of SegGPT foundation model to address these issues, especially in light of the limitations seen with traditional deep learning methods.

Even considering that the used models helped us answer the main questions of this thesis, our findings make us glimpse future improvements that can be investigated. Below, we outline some of these potential investigative paths.

**First study:** In the cross-species generalization task, we can extend segmentation to other renal histological structures and increase the number of samples in the WSI dataset. Additionally, applying foundation models for intra-dataset evaluations and mouse-to-human generalization offers a valuable opportunity. We also plan to acquire more mice samples stained with PAS and PAMS to further explore cross-species generalization. Other species not investigated in this study may also be considered to evaluate the generalization potential.

**Second study:** For the borderless glomeruli task, we aim to investigate additional foundation models specifically focused on glomeruli affected by global and segmental sclerosis. Furthermore, we intend to evaluate the percentage of affected areas within each segmented glomerulus exhibiting sclerosis. Additionally, we seek to explore the

application of foundation models for segmenting other glomerular lesions and renal histological structures.

In our comprehensive studies on medical imaging, we developed deep learning and foundation model methods designed to segment glomeruli in high-resolution histopathological images, even with limited training data. The aim was to develop an automated framework to assist renal pathologists in analyzing WSIs, particularly for glomeruli with presence, partial, and absent borders.





## REFERENCES

- ADAMS, R.; BISCHOF, L. Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, v. 16, n. 6, p. 641–647, 1994.
- AGARWAL, S.; SETHI, S.; DINDA, A. Basics of kidney biopsy: A nephrologist's perspective. *Indian journal of nephrology*, Wolters Kluwer–Medknow Publications, v. 23, n. 4, p. 243, 2013.
- ALTINI, N. et al. Semantic segmentation framework for glomeruli detection and classification in kidney histological sections. *Electronics*, Multidisciplinary Digital Publishing Institute, v. 9, n. 3, p. 503, 2020.
- ALTINI, N. et al. Performance and limitations of a supervised deep learning approach for the histopathological oxford classification of glomeruli with iga nephropathy. *Computer Methods and Programs in Biomedicine*, p. 107814, 2023.
- ALTURKISTANI, H. A.; TASHKANDI, F. M.; MOHAMMEDSALEH, Z. M. Histological stains: a literature review and case study. *Global journal of health science*, Canadian Center of Science and Education, v. 8, n. 3, p. 72, 2016.
- ASADZADEH, S. et al. Renal structural image processing techniques: a systematic review. *Renal failure*, Taylor & Francis, v. 41, n. 1, p. 57–68, 2019.
- ATHANAZIO, D. A. et al. O ensino de patologia nas escolas médicas está em crise? uma revisão sobre a experiência internacional. *Revista Brasileira de Educação Médica*, SciELO Brasil, v. 33, p. 49–54, 2009.
- BANKMAN, I. *Handbook of medical imaging: processing and analysis management*. [S.l.]: Academic press, 2000.
- BARISONI, L. et al. Digital pathology evaluation in the multicenter nephrotic syndrome study network (neptune). *Clinical Journal of the American Society of Nephrology*, Am Soc Nephrol, v. 8, n. 8, p. 1449–1459, 2013.
- BAYRAMOGLU, N.; KANNALA, J.; HEIKKILÄ, J. Deep learning for magnification independent breast cancer histopathology image classification. In: *IEEE. 2016 23rd International conference on pattern recognition (ICPR)*. [S.l.], 2016. p. 2440–2445.
- BEL, T. de et al. Automatic segmentation of histopathological slides of renal tissue using deep learning. In: *INTERNATIONAL SOCIETY FOR OPTICS AND*

- PHOTONICS. *Medical Imaging 2018: Digital Pathology*. [S.l.], 2018. v. 10581, p. 1058112.
- BERMAN, M.; TRIKI, A. R.; BLASCHKO, M. B. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 4413–4421.
- BOUTELDJA, N. et al. Deep learning-based segmentation and quantification in experimental kidney histopathology. *Journal of the American Society of Nephrology, Am Soc Nephrol*, v. 32, n. 1, p. 52–68, 2021.
- BUENO, G. et al. Glomerulosclerosis identification in whole slide images using semantic segmentation. *Computer methods and programs in biomedicine*, Elsevier, v. 184, p. 105273, 2020.
- CAO, H. et al. Swin-unet: Unet-like pure transformer for medical image segmentation. In: SPRINGER. *European conference on computer vision*. [S.l.], 2022. p. 205–218.
- CARNEIRO, G. et al. Review of deep learning methods in mammography, cardiovascular, and microscopy image analysis. In: *Deep Learning and Convolutional Neural Networks for Medical Image Computing*. [S.l.]: Springer, 2017. p. 11–32.
- CATHRO, H. P.; SHEN, S. S.; TRUONG, L. D. Diagnostic histochemistry in medical diseases of the kidney. In: ELSEVIER. *Seminars in diagnostic pathology*. [S.l.], 2018. v. 35, p. 360–369.
- CHAGAS, P. et al. Uncertainty-aware membranous nephropathy classification: A monte-carlo dropout approach to detect how certain is the model. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Taylor & Francis, p. 1–11, 2022.
- CHAI, J. et al. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, Elsevier, v. 6, p. 100134, 2021.
- CHANG, A. et al. A position paper on standardizing the nonneoplastic kidney biopsy report. *Human pathology*, Elsevier, v. 43, n. 8, p. 1192–1196, 2012.
- CHATER, S. et al. Learning from mouse ct-scan brain images to detect mra-tof human vasculatures. In: IEEE. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. [S.l.], 2021. p. 2830–2834.
- CHEN, L.-C. et al. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

- CHEN, L.-C. et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European conference on computer vision (ECCV)*. [S.l.: s.n.], 2018. p. 801–818.
- CHEN, R. J. et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, Nature Publishing Group US New York, v. 30, n. 3, p. 850–862, 2024.
- CHEN, Z.; SUN, Q. Weakly-supervised semantic segmentation with image-level labels: from traditional models to foundation models. *ACM Computing Surveys*, ACM New York, NY, 2023.
- DAVIS, R. C. et al. Deep learning segmentation of glomeruli on kidney donor frozen sections. *medRxiv*, Cold Spring Harbor Laboratory Press, p. 1–27, 2021.
- Della Mea, V. et al. User attitudes in analyzing digital slides in a quality control test bed: A preliminary study. *Computer Methods and Programs in Biomedicine*, v. 82, n. 2, p. 177–186, 2006.
- DOS-SANTOS, W. L. C. et al. Current distribution pattern of biopsy-proven glomerular disease in salvador, brazil, 40 years after an initial assessment. *Brazilian Journal of Nephrology*, SciELO Brasil, v. 39, p. 376–383, 2017.
- FAN, T. et al. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, IEEE, v. 8, p. 179656–179665, 2020.
- FARRIS, A. B. et al. Whole slide imaging for analytical anatomic pathology and telepathology: practical applications today, promises, and perils. *Archives of pathology & laboratory medicine*, the College of American Pathologists, v. 141, n. 4, p. 542–550, 2017.
- GADERMAYR, M. et al. Cnn cascades for segmenting sparse objects in gigapixel whole slide images. *Computerized Medical Imaging and Graphics*, Elsevier, v. 71, p. 40–48, 2019.
- GALLEGO, J. et al. A u-net based framework to quantify glomerulosclerosis in digitized pas and h&e stained human tissues. *Computerized Medical Imaging and Graphics*, Elsevier, v. 89, p. 101865, 2021.
- GESUALDO, L. et al. The italian experience of the national registry of renal biopsies. *Kidney International*, Elsevier, v. 66, n. 3, p. 890–894, 2004.
- GINLEY, B. et al. Neural network segmentation of interstitial fibrosis, tubular atrophy, and glomerulosclerosis in renal biopsies. *arXiv preprint arXiv:2002.12868*, 2020.
- GINLEY, B. et al. Computational segmentation and classification of diabetic glomerulosclerosis. *Journal of the American Society of Nephrology*, Am Soc Nephrol, v. 30, n. 10, p. 1953–1967, 2019.

- GONZALEZ, R. C. *Digital image processing*. [S.l.]: Pearson education india, 2009.
- HAAS, M. et al. Consensus definitions for glomerular lesions by light and electron microscopy: recommendations from a working group of the renal pathology society. *Kidney International*, Elsevier, v. 98, n. 5, p. 1120–1134, 2020.
- HERMSEN, M. et al. Deep learning–based histopathologic assessment of kidney tissue. *Journal of the American Society of Nephrology*, Am Soc Nephrol, v. 30, n. 10, p. 1968–1979, 2019.
- HOGAN, J. J.; MOCANU, M.; BERNS, J. S. The native kidney biopsy: update and evidence for best practice. *Clinical Journal of the American Society of Nephrology*, Am Soc Nephrol, v. 11, n. 2, p. 354–362, 2016.
- HOSSAIN, M. M.; SAHARKHIZ, N.; KONOFAGOU, E. E. Feasibility of harmonic motion imaging using a single transducer: In vivo imaging of breast cancer in a mouse model and human subjects. *IEEE Transactions on Medical Imaging*, IEEE, v. 40, n. 5, p. 1390–1404, 2021.
- HOWARD, A. et al. *HuBMAP - Hacking the Kidney: Identify glomeruli in human kidney tissue images*. 2020. Kaggle.
- HUANG, H. et al. Unet 3+: A full-scale connected unet for medical image segmentation. In: IEEE. *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.], 2020. p. 1055–1059.
- ING, N. et al. Semantic segmentation for prostate cancer grading by convolutional neural networks. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2018: Digital Pathology*. [S.l.], 2018. v. 10581, p. 105811B.
- JAYAPANDIAN, C. P. et al. Development and evaluation of deep learning–based segmentation of histologic structures in the kidney cortex with multiple histologic stains. *Kidney International*, v. 99, n. 1, p. 86–101, 2021.
- JHA, A. et al. Instance segmentation for whole slide imaging: end-to-end or detect-then-segment. *Journal of Medical Imaging*, International Society for Optics and Photonics, v. 8, n. 1, p. 014001, 2021.
- JIANG, L. et al. A deep learning-based approach for glomeruli instance segmentation from multistained renal biopsy pathologic images. *The American Journal of Pathology*, Elsevier, v. 191, n. 8, p. 1431–1441, 2021.
- KANNAN, S. et al. Segmentation of glomeruli within trichrome images using deep learning. *Kidney International Reports*, Elsevier, v. 4, n. 7, p. 955–962, 2019.
- KAWAZOE, Y. et al. Computational pipeline for glomerular segmentation and association of the quantified regions with prognosis of kidney function in iga nephropathy. *Diagnostics*, MDPI, v. 12, n. 12, p. 2955, 2022.

- KIM, H.-R. et al. Comparative gene expression profiles of intestinal transporters in mice, rats and humans. *Pharmacological research*, Elsevier, v. 56, n. 3, p. 224–236, 2007.
- KOMURA, D.; ISHIKAWA, S. Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*, Elsevier, v. 16, p. 34–42, 2018.
- LEVEY, A. S. et al. National kidney foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Annals of internal medicine*, Am Coll Physicians, v. 139, n. 2, p. 137–147, 2003.
- LI, X. et al. Deep learning segmentation of glomeruli on kidney donor frozen sections. *Journal of Medical Imaging*, Society of Photo-Optical Instrumentation Engineers, v. 8, n. 6, p. 067501–067501, 2021.
- LIU, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2021. p. 10012–10022.
- LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. p. 3431–3440.
- LUO, H.; CHANGDONG, Y.; SELVAN, R. Hybrid ladder transformers with efficient parallel-cross attention for medical image segmentation. In: PMLR. *International Conference on Medical Imaging with Deep Learning*. [S.l.], 2022. p. 808–819.
- LUTNICK, B. et al. An integrated iterative annotation technique for easing neural network training in medical image analysis. *Nature machine intelligence*, Nature Publishing Group, v. 1, n. 2, p. 112–119, 2019.
- LYON, H.; HOROBIN, R. Standardization and standards for dyes and stains used in biology and medicine. *Biotechnic & Histochemistry*, Taylor & Francis, v. 82, n. 1, p. 1–11, 2007.
- LYON, H. O. et al. Standardization of reagents and methods used in cytological and histological practice with emphasis on dyes, stains and chromogenic reagents. *The Histochemical Journal*, Springer, v. 26, n. 7, p. 533–544, 1994.
- MARSH, J. N. et al. Deep learning global glomerulosclerosis in transplant kidney frozen sections. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 12, p. 2718–2728, 2018.
- NIAZI, M. K. K.; PARWANI, A. V.; GURCAN, M. N. Digital pathology and artificial intelligence. *The lancet oncology*, Elsevier, v. 20, n. 5, p. e253–e261, 2019.

- PASZKE, A. et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, v. 32, 2019.
- POLITO, M. G.; MOURA, L. A. R. D.; KIRSZTAJN, G. M. An overview on frequency of renal biopsy diagnosis in brazil: clinical and pathological patterns based on 9617 native kidney biopsies. *Nephrology dialysis transplantation*, Oxford University Press, v. 25, n. 2, p. 490–496, 2010.
- RAO, S.; SETTY, S.; KATIKIREDDI, R. S. Morphometric study of human adult cadaveric kidneys-research article. *International journal of current research and review*, v. 5, p. 109–115, 2013.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical Image Computing and Computer-assisted Intervention*. [S.l.], 2015. p. 234–241.
- RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, Springer, v. 115, n. 3, p. 211–252, 2015.
- SAIKIA, F. N. et al. Mlp-unet: Glomerulus segmentation. *IEEE Access*, IEEE, v. 11, p. 53034–53047, 2023.
- SANTOS, J. D. et al. Classificação de imagens de biópsias renais com glomeruloesclerose segmentar e focal ou com lesões mínimas utilizando transfer learning em cnn. In: SBC. *Anais do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*. [S.l.], 2019. p. 82–93.
- SCHWARTZ, M. et al. *Pathology of the kidney*. [S.l.]: New York: Lippincott Williams & Wilkins, 2007.
- SHAMSHAD, F. et al. Transformers in medical imaging: A survey. *Medical Image Analysis*, Elsevier, v. 88, p. 102802, 2023.
- SILVA, J. et al. Boundary-aware glomerulus segmentation: toward one-to-many stain generalization. *Computerized Medical Imaging and Graphics*, Elsevier, v. 100, p. 102104, 2022.
- SILVA, J. et al. Boundary-aware glomerulus segmentation: Toward one-to-many stain generalization. *Computerized Medical Imaging and Graphics*, v. 100, p. 102104, 2022.
- SIMON, O. et al. Multi-radial lbp features as a tool for rapid glomerular detection and assessment in whole slide histopathology images. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 2032, 2018.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

- SMITH, A. L.; CORROW, D. J. Modifications to husbandry and housing conditions of laboratory rodents for improved well-being. *ILAR journal*, Institute for Laboratory Animal Research, v. 46, n. 2, p. 140–147, 2005.
- SOUZA, L. et al. How feasible is it to segment human glomerulus with a model trained on mouse histology images? In: SBC. *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. [S.l.], 2022. p. 449–460.
- SOUZA, L. et al. Mouse-to-human transfer learning for glomerulus segmentation. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Taylor & Francis, v. 11, n. 5, p. 1996–2005, 2023.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. *International Conference on Machine Learning*. [S.l.], 2019. p. 6105–6114.
- VALANARASU, J. M. J. et al. Medical transformer: Gated axial-attention for medical image segmentation. In: SPRINGER. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*. [S.l.], 2021. p. 36–46.
- WALKER, P. D.; CAVALLO, T.; BONSI, S. M. Practice guidelines for the renal biopsy. *Modern Pathology*, Nature Publishing Group, v. 17, n. 12, p. 1555, 2004.
- WANG, X. et al. A mouse 5/6th nephrectomy model that induces experimental uremic cardiomyopathy. *Journal of visualized experiments: JoVE*, MyJoVE Corporation, n. 129, 2017.
- WANG, X. et al. Images speak in images: A generalist painter for in-context visual learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2023. p. 6830–6839.
- WANG, X. et al. Seggpt: Towards segmenting everything in context. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.: s.n.], 2023. p. 1130–1140.
- YI-DE, M.; QING, L.; ZHI-BAI, Q. Automated image segmentation using improved pcnn model based on cross-entropy. In: IEEE. *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004*. [S.l.], 2004. p. 743–746.
- ZHANG, Z.; LIU, Q.; WANG, Y. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, IEEE, v. 15, n. 5, p. 749–753, 2018.
- ZHANG, Z.; ZHANG, W. Pyramid medical transformer for medical image segmentation. *arXiv preprint arXiv:2104.14702*, 2021.

ZHOU, T. et al. Image segmentation in foundation model era: A survey. *arXiv preprint arXiv:2408.12957*, 2024.