Computerized Medical Imaging and Graphics 126 (2025) 102658

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Multistain multicompartment automatic segmentation in renal biopsies with thrombotic microangiopathies and other vasculopathies

Nicola Altini [a,1], Michela Prunella [a,1], Surya V. Seshan [b], Savino Sciascia [c], Antonella Barreca [d], Alessandro Del Gobbo [e], Stefan Porubsky [f], Hien Van Nguyen [g], Claudia Delprete [a], Berardino Prencipe [a], Deján Dobi [h], Daan P.C. van Doorn [i,j], Sjoerd A.M.E.G. Timmermans [i,j], Pieter van Paassen [i,j], Vitoantonio Bevilacqua [a,*,2], Jan Ulrich Becker [k,2]

[a] Department of Electrical and Information Engineering, Polytechnic University of Bari, Bari, Italy
[b] Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA
[c] Department of Clinical and Biological Sciences, University of Turin, Turin, Italy
[d] Division of Pathology, Città Della Salute e Della Scienza Hospital, University of Turin, Turin, Italy
[e] Division of Pathology, Fondazione IRCCS Ca' Granda – Ospedale Maggiore Policlinico, Milan, Italy
[f] Department of Pathology, University Medical Center, Johannes Gutenberg University Mainz, Mainz, Germany
[g] Department of Electrical and Computer Engineering, University of Houston, Houston, USA
[h] Department of Pathology, Forensic and Insurance Medicine, Semmelweis University, Budapest, Hungary
[i] Expert Center for Immune-mediated Kidney Diseases and Vasculitis, Maastricht University Medical Center, Maastricht, Netherlands
[j] Department of Biochemistry, Cardiovascular Research Institute Maastricht, Maastricht, Netherlands
[k] Institute of Pathology, University Hospital of Cologne, Cologne, Germany

ARTICLE INFO

Keywords:
Nephropathology
Segmentation
Deep Learning
Thrombotic Microangiopathy

ABSTRACT

Automatic tissue segmentation is a necessary step for the bulk analysis of whole slide images (WSIs) from paraffin histology sections in kidney biopsies. However, existing models often fail to generalize across the main nephropathological staining methods and to capture the severe morphological distortions in arteries, arterioles, and glomeruli common in thrombotic microangiopathy (TMA) or other vasculopathies. Therefore, we developed an automatic multi-staining segmentation pipeline covering six key compartments: Artery, Arteriole, Glomerulus, Cortex, Medulla, and Capsule/Other. This framework enables downstream tasks such as counting and labeling at instance-, WSI- or biopsy-level. Biopsies (n = 158) from seven centers: Cologne, Turin, Milan, Weill-Cornell, Mainz, Maastricht, Budapest, were classified by expert nephropathologists into TMA (n = 87) or Mimickers (n = 71). Ground truth expert segmentation masks were provided for all compartments, and expert binary TMA classification labels for Glomerulus, Artery, Arteriole. The biopsies were divided into training (n = 79), validation (n = 26), and test (n = 53) subsets. We benchmarked six deep learning models for semantic segmentation (U-Net, FPN, DeepLabV3+, Mask2Former, SegFormer, SegNeXt) and five models for classification (ResNet-34, DenseNet-121, EfficientNet-v2-S, ConvNeXt-Small, Swin-v2-B). We obtained robust segmentation results across all compartments. On the test set, the best models achieved Dice coefficients of 0.903 (Cortex), 0.834 (Medulla), 0.816 (Capsule/Other), 0.922 (Glomerulus), 0.822 (Artery), and 0.553 (Arteriole). The best classification models achieved Accuracy of 0.724 and 0.841 for Glomerulus and Artery plus Arteriole compartments, respectively. Furthermore, we release NePathTK (NephroPathology Toolkit), a powerful open-source end-to-end pipeline integrated with QuPath, enabling accurate segmentation for decision support in nephropathology and large-scale analysis of kidney biopsies.

* Corresponding author.
E-mail address: vitoantonio.bevilacqua@poliba.it (V. Bevilacqua).
[1] Joint first author
[2] Joint senior author

https://doi.org/10.1016/j.compmedimag.2025.102658
Received 3 June 2025; Received in revised form 17 September 2025; Accepted 15 October 2025
Available online 22 October 2025
0895-6111/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Table 1
Related works concerning the segmentation of histological compartments from renal whole slide images. Abbreviation for Staining: PAS (periodic acid-Schiff), PAM (periodic acid-silver methenamine), H&E (hematoxylin and eosin), TRI (trichrome), PAS-H (PAS with hematoxylin counterstain), CR (Congo red), SR (Sirius red), JH&E (Jones H&E).

Reference	Study Objective	Histological Compartments	Staining	Samples Origin
(Bueno et al., 2020)	Quantify glomerulosclerosis	Glomeruli	PAS	human kidney biopsies
(Altini et al., 2020b,a)	Assess pre-transplant kidney functionality in the Karpinski score	Glomeruli	PAS	human donor kidney biopsies
(Salvi et al., 2021)	Assess pre-transplant kidney functionality in the Karpinski score	Glomeruli, tubules	PAS	human donor kidney biopsies
(Jiang et al., 2021)	Classification of glomeruli into sclerosed, other disease or healthy	Glomeruli	PAS, PAM, TRI	human kidney biopsies
(Lutnick et al., 2022)	Segmentation tool for nephropathologists	Glomeruli, Arteries, Arterioles, IFTA	H&E, PAS, PAS-H, PAM, TRI, CR	human kidney biopsies, transplant biopsies, murine renal tissues
(Bouteldja et al., 2021)	Increase reproducibility and efficiency of histopathological analyses in preclinical animal models	Glomerular tuft, glomerulus including Bowman's capsule, tubules, arteries, arterial lumina, and veins	PAS	healthy mice and murine diseased model and other species
(Jayapandian et al., 2021)	Quantification of histologic primitives on renal parenchyma of minimal change disease	Glomerular tuft and unit, proximal/distal tubule, arteries and afferent arterioles, and peritubular capillaries	H&E, PAS, PAM, TRI	human kidney biopsies and nephrectomies
(Vasiljević et al., 2021)	Learning stain-invariant structure descriptors	Glomeruli	H&E, PAS, JH&E, CD68, SR, CD34, CD3	human kidney nephrectomies
(Silva et al., 2022)	Segmenting glomeruli in different stained WSIs	Glomeruli (tuft, capsule)	H&E, PAS, PAM, TRI	human kidney biopsies
(Hölscher et al., 2023)	Prognosis prediction in IgA-nephropathy	Glomeruli (tuft and unit), tubules, arteries, arterial lumen	PAS	human kidney biopsies and nephrectomies
(Altini et al., 2023)	Oxford Classification for IgA nephropathy	Glomeruli	PAS	human kidney biopsies
(Vasiljević et al., 2023)	Stain transfer between multiple stainings, stain normalization and stain invariant segmentation	Glomeruli	PAS, JH&E, SR, CD34, CD68	human kidney nephrectomies

1. Introduction

Precision nephropathology requires reproducible and comprehensive evaluation of the morphological changes in whole slide images (WSIs) from paraffin sections. Previous works showed that machine learning (ML) and computer vision (CV) approaches are the only realistic option to correlate nephropathological imaging data with clinically useful diagnostic, prognostic and theranostic endpoints in big multicentric datasets (Becker et al., 2020). Moreover, the increasingly ubiquitous slide scanners and ever more powerful graphic processing units make this approach of digital nephropathology an increasingly accurate method to develop cost-efficient and, crucially, also evidence-based clinical decision support tools of tomorrow.

Nephropathology is a special discipline in histopathology, because it uses a wide array of paraffin stainings, including trichrome (TRI), periodic-acid Schiff (PAS) and periodic acid-silver methenamine (PAM) stainings in addition to the standard histopathological staining of hematoxylin and eosin (HE). These four standard stainings are then used by the nephropathologist in a complementary fashion to examine the main histopathological compartments: arteries, arterioles, glomeruli, cortex (tubulointerstitium) and medulla for a set of consensus descriptors (Haas et al., 2020). Together with results from immunostaining (usually immunofluorescence) for immunoglobulins and complement split products and transmission electron micrographs, these descriptors are then integrated by the nephropathologist with further clinical information to provide diagnostic and prognostic information, thus helping with treatment decisions.

One of the first applications of ML and CV in nephropathology were segmentation models. Beyond their use for simple tasks, such as counting glomeruli in a biopsy, these segmentation models were useful as a first step before feature extraction of mainly glomeruli (Ginley et al., 2019) but also other compartments (Hölscher et al., 2023). More recent works used either fully supervised approaches, where manually or automatically segmented glomerular compartments are used to replicate expert nephropathology scores at instance-level (Altini et al., 2023), or weakly supervised approaches, where models are trained to infer a biopsy-level diagnostic label from bags of glomeruli (Zare et al., 2025).

With the surge of advanced image analysis functionalities embodied in deep learning (DL) frameworks (Berloco et al., 2025), augmented workflows have been refined to assist nephropathologists in decision-making through the objective scoring of a large volume of images, easing the specialists' workload while enhancing precision and reproducibility.

In this study, our aim is to engineer a fully automated pipeline for the precise bulk segmentation of kidney WSIs encompassing a wide variety of cases with thrombotic microangiopathies (TMAs) or other vasculopathies (that we refer to as Mimickers), collected in an international multicenter collaboration, spanning seven centers: Cologne (Germany), Weill-Cornell (New York, NY, USA), Turin (Italy), Milan (Italy), Mainz (Germany), Maastricht (Netherlands), Budapest (Hungary). This unprecedented expert curated WSIs dataset, encompassing the accepted tetrad of stainings deemed useful for diagnosing TMA, allowed for the simultaneous characterization of multiple renal histopathological compartments with high diagnostic and prognostic value. We designed our study as a many-to-many segmentation (Silva et al., 2022), by mixing different stains in the sets of training and testing, to adequately express the mechanisms of TMA and its mimickers' and their phenotypic traits. Remarkably, a precise and effective segmentation of histopathological structures as disease indicators, is required for pathomic data mining, as showed in the morphometric analysis FLASH conducted on other relevant kidney diseases (Hölscher et al., 2023). Furthermore, they could also be exploited for further downstream applications of instance-level classification or multiple instance-learning with bags of compartment crops.

Our contributions are summarized as follows:

1) We considered a segmentation task for nephropathology images that is more comprehensive than previous ones, encompassing both the "coarse" structures of the kidney, consisting of cortex, medulla, and

capsule/other, and the "fine" structures, i.e., glomeruli, arteries, and arterioles.
2) We conceived a classification task for TMA classification at instance-level, for the aforementioned "fine" structures. Two different configurations were considered: one for glomerular instances, and the other one grouping arteries and arterioles.
3) A thorough comparison of convolutional neural networks (CNN)-based and Transformer-based semantic segmentation and instance-level TMA classification models, along with hybrid CNN with Transformers blocks and multiscale attention models on our internationally collected cohorts.
4) The proposed framework was developed and validated through an international multi-center, multi-stained, and multi-compartment annotated kidney slide dataset including severe histopathological alterations.
5) The development of a tool that can be leveraged by nephrologists, pathologists, computer scientists and other interested practitioners, NePathTK (NephroPathology Toolkit), for the automatic segmentation and quantification of relevant tissue structures from kidney biopsies, also in the presence of severely damaged histological traits as in the case of TMA. Furthermore, the tool can be used to recognize, at instance-level, the occurrence of TMA, hence working as an effective screening tool for this subtle and rare pathology.

2. Related Works

In the related literature, computerized tasks involved the detection and counting of histological structures required for subsequent extraction of informative features using DL (Bueno et al., 2020; Hölscher et al., 2023). The implemented DL pipelines consisted either of one-step approaches, involving CNNs and/or transformer-based encoders for simultaneous segmentation or detection and classification, or in two-step approaches, where classification is dealt only after the regions of interest have been segmented (Bueno et al., 2020; Jiang et al., 2021). These works demonstrated the capability of automated systems to objectively assess the kidney functionality, as well as for increasing the reproducibility of existing scoring systems, such as the Karpinski score (Altini et al., 2020b,a; Salvi et al., 2021) for evaluating transplant eligibility, MESC (Altini et al., 2023) and MEST-C (Jaugey et al., 2023) Oxford classification in IgA nephropathy, Banff classification for transplant biopsies (Kers et al., 2022), classification of Antibody Mediated Rejection (AMR) (Zare et al., 2025), and pathological classification of diabetic nephropathy (Ginley et al., 2019), among the leading applications.

Most existing works consider the segmentation of glomeruli and tubules on PAS-stained WSIs. Glomeruli were classified into sclerotic, healthy, or other lesions (Jiang et al., 2021), mesangioproliferation, endocapillary hypercellularity, crescents, segmental sclerosis (Altini et al., 2023; Jaugey et al., 2023), while tubules in atrophic (Salvi et al., 2021) or involved into interstitial fibrosis and tubular atrophy (IFTA) (Lutnick et al., 2022), respectively. Existing DL frameworks approached the localization and segmentation of kidney structures through tailored networks, elaborating only a single or a couple of stains for training and evaluation. This approach is commonly observed in most of the related works summarized in Table 1, which poses a domain-shift challenge when reproducing experiments in different settings.

Vasiljević et al. (2021), developed a Generative Adversarial Network-based approach able to learn stain-invariant representations of glomeruli, by transforming a source domain consisting in PAS-stained crops to two target domains: Jones HE and CD34. Silva et al. (2022), developed a glomeruli segmentation network to approach a single-to-multi stain learning, in which low-level glomeruli boundary information was passed in higher semantic layers and then fused with an attention map; their DS-FNet was trained only on PAS-stained WSIs achieving 0.92 Dice coefficient on PAS-stained test set, and average 0.84 Dice when tested in PAM (123 WSIs), HE (157 WSIs), and TRI (137 WSIs) in the NEPTUNE dataset. Jayapandian et al. (2021), assessed the influences of the number of training samples, magnification factor and of the used stain on the segmentation performance of glomeruli, tubules, arteries/arterioles and peritubular capillaries. However, only cases of minimal change disease, which is characterized by subtle or minimal changes in kidney tissue making it very similar to normal renal parenchyma, were included in the study.

The compartment of arterioles is notoriously difficult to detect for expert nephropathologists. Among the previously published studies, Lutnick et al. (2022), addressed the artery/arteriole segmentation task in diseased nephropathological biopsies by fine tuning of DeepLabV3+ (Chen et al., 2018) on 226 WSIs and testing it on 58 WSIs; however, the relative paucity of severe vasculopathies in the dataset might have resulted in an inferior training set.

There is clearly a need for the representation of not only all four main nephropathological stainings, but also for severe vasculopathies and their morphological alterations in training datasets for segmentation models. The TMAs encompass a wide histomorphological spectrum of alterations coming with different possible etiologies, which need to be adequately represented in future DL frameworks in order to improve diagnostic accuracy over other conditions that share some common features.

3. Materials and methods

3.1. Dataset

All n = 158 biopsies were diagnosed by experienced nephropathologists in Cologne (JUB), at Weill-Cornell (SVS), in Turin (AB), in Milan (ADG), in Mainz (SP), in Maastricht (SAMEGT and PVP), and in Budapest (DD).

The TMA cohort consisted of n = 87 biopsies with a histopathological diagnosis of TMAs (n = 33 from Cologne, n = 5 from Turin, n = 10 from Weill-Cornell, n = 5 from Milan, n = 15 from Mainz, n = 10 from Maastricht, n = 9 from Budapest). Inclusion criteria for TMAs included microthrombi, split glomerular basement membranes without any other apparent cause, dysoria of arteries and arterioles, with intra- or extramural leakage of blood constituents, foam cell, myxoid and hypoelastic intimal changes as well as onion-skin transformation and obliteration of arteries and arterioles. The aetiology of TMA included arterial hypertension, bacterial infection, systemic sclerosis, anti-phospholipid antibody disease, chemotherapy/Avastin treatment, pregnancy-associated TMA and atypical hemolytic uremic syndrome of unknown aetiology.

The second cohort of n = 71 Mimickers consisted of severe renal vasculopathies (n = 36 from Cologne, n = 5 from Turin, n = 6 from Weill-Cornell, n = 5 from Milan, n = 9 from Mainz, n = 10 from Maastricht). Mimickers were biopsies with a diagnosis showing similar histological features to TMA, often resulting in severe distortions of compartment morphology, particularly in arteries, arterioles and/or glomeruli. These diseases included severe benign nephrosclerosis (n = 25), anti-neutrophil cytoplasmic antibody (ANCA) associated glomerulonephritis with leukocytoclastic arteritis (n = 18), cryoglobulinemic glomerulonephritis/arteritis (n = 9), Bevacizumab-associated obliterative glomerulopathy (Pfister et al., 2018) (n = 4), collapsing focal and segmental glomerulosclerosis (n = 4), cholesterol emboli (n = 3), light chain deposition disease with arteriolopathy (n = 2), hydrophilic polymer nephropathy (n = 1), Castleman's disease (n = 1), AA amyloidosis (n = 1), Waldenström's macroglobulinemia (n = 1), intravascular large B cell lymphoma (n = 1), C3 glomerulonephritis (n = 1).

The training set for the segmentation model consisted of n = 79 biopsies (n = 51 from Cologne, n = 8 from Turin, n = 12 from Weill-Cornell, n = 8 from Milan). The validation set consisted of n = 26 biopsies (n = 18 from Cologne, n = 2 from Turin, n = 4 from Weill-Cornell, n = 2 from Milan). The test set for the segmentation model consisted of n = 53 biopsies (n = 24 from Mainz, n = 20 from Maastricht, n = 9 from

Table 2
Stainings composition for the cohorts. Details of the number of WSIs for each of the main four nephropathological stains for each institution participating in the study: Cologne, Milan, Turin, Weill-Cornell, Mainz, Maastricht, Budapest.

Institution	HE	PAS	PAM	TRI
Cologne	103	170	105	172
Milan	10	8	9	12
Turin	0	10	8	10
Weill-Cornell	16	17	17	17
Mainz	21	24	0	0
Maastricht	19	20	20	19
Budapest	6	4	2	6

Budapest). We included HE, PAS, PAM, and TRI stainings in this project. TRI stainings include Masson's (MT), Verhoeff's Van Gieson (VVG), acid-fuchsin orange G (AFOG), Masson-Goldner (MASG), and Masson-Goldner-Elastica (MGE).

The biopsies from Cologne and Turin were scanned with a Hamamatsu Scanner (Hamamatsu Photonics, Herrsching am Ammersee, Germany). The biopsies from Weill-Cornell were scanned with an Aperio GT 450 scanner (Leica Biosystems, Deer Park, IL, USA). The biopsies from Milan were scanned with an Aperio scanner (Leica Biosystems, Deer Park, IL, USA). The biopsies from Mainz were scanned with a Hamamatsu Scanner (Hamamatsu Photonics, Herrsching am Ammersee, Germany). The biopsies from Maastricht were scanned with a Pannoramic 1000 Digital Scanner (3DHISTECH, Budapest, Hungary). The biopsies from Budapest were scanned with a Pannoramic Scanner (3DHISTECH, Budapest, Hungary). All WSIs were obtained with either a $\times$ 20 or a $\times$ 40 objective.

The number of WSIs for each of the four stainings (HE, PAS, PAM, TRI) is reported in Table 2.

3.2. Expert annotation

Expert annotations were provided for both the segmentation of the six key compartments (Glomerulus, Artery, Arteriole, Cortex, Medulla, Capsule/Other), as described in Section 3.2.1, and for the instance-level classification of TMA from the "fine" compartments (Glomerulus, Artery, Arteriole), as described in Section 3.2.2.

3.2.1. Segmentation

All expert annotations (manual segmentations) were performed under QuPath (Bankhead et al., 2017) on a Wacom Cintiq Pro 16-inch or 24-inch pen display (Wacom, Düsseldorf, Germany) by an experienced nephropathologist (JUB). Object classes included Glomerulus, Arteriole, Artery, Cortex, Medulla, Capsule/Other, and Ignore. For arterial and arteriolar annotations, we included the walls up to the outer border of the media as well as the lumina; all luminal areas in incompletely captured, "opened" segments were included. Arteries were distinguished from Arterioles by the number of smooth muscle cell layers, with at least 2 over the entire circumference defining an Artery (Roufosse and Simmonds, 2018). Glomeruli were annotated including the tuft, Bowman's space and the perimeter of Bowman's capsule. Empty Bowman's capsules were disregarded for glomerular annotation. Dislodged glomerular tufts were included; tufts, if still attached to the capsule, were included with the capsule segment; if completely detached, under inclusion of the entire tuft only. Ignore was used for artefacts such as out-of-focus areas, tissue folds or dirt.

Cortex was differentiated from Medulla by the characteristic tubular morphology and by the presence of glomeruli in the former and the characteristic parallel alignment of tubules/ducts and the absence of arteries in the latter; Capsule/Other contained renal capsule and other non-renal tissue elements such as connective tissue, skeletal muscle and other incidental tissue/organ elements contained in the biopsy.

In the training set, n = 2456, n = 15,580, and n = 10,978 objects were annotated for the three classes: Artery, Arteriole, Glomerulus, and n = 2377, n = 762, and n = 2069 objects were annotated for Cortex, Medulla, Capsule/Other. In the validation set, n = 542, n = 4058, and n = 3726 objects were annotated for the three classes: Artery, Arteriole, Glomerulus, and n = 548, n = 158, and n = 722 objects were annotated for Cortex, Medulla, Capsule/Other. In the test set, n = 589, n = 4239, and n = 3296 objects were annotated for the three classes: Artery, Arteriole, Glomerulus, and n = 552, n = 278, and n = 332 objects were annotated for Cortex, Medulla, Capsule/Other.

The details of how the objects are subdivided among the seven institutions participating in this study are reported in Table 3.

Exemplary PAS images from the seven institutions are shown in Fig. 1. PAS staining was chosen for this figure since it is the only stain which is available from all centers. Examples of the four main nephropathological stainings (i.e., HE, PAS, PAM, TRI) are shown in Fig. 2.

3.2.2. Classification

Consensus nephropathology definitions for TMAs have only been developed for kidney transplant biopsies (Afrouzian et al., 2023). In the absence of a consensus definition for native biopsies, all nephropathologists involved in this study (SS, AB, DD, SAMEGT, JUB) agreed on a working definition for a set of binary (absent or present) histopathological descriptors for the three compartments Artery, Arteriole, Glomerulus.

Considering that TMA lesions would appear similar across Artery and Arteriole compartments, if compared to the Glomerulus compartment, the following definitions were used:

- Artery/Arteriole: Thrombus (platelet, fibrin or mixed), foam cell in intima, hypoelastotic intimal fibrosis, onion skin fibrosis, wall edema, myxoid intimal fibrosis, fibrin insudation, erythrocyte insudation, wall necrosis, endothelial denudation, endotheliosis.
- Glomerulus: microthrombus (platelet, fibrin, mixed), GBM splitting, microaneurysm, glomerular collapse, glomerular infarct, endotheliosis, mesangiolysis.

The presence of one or a combination of these descriptors in any compartment was considered indicative of TMA.

A total of n = 9591 glomerular annotations, n = 12,624 arteriole annotations, and n = 1918 arteries annotations were extracted from 86 biopsies with TMA, using QuPath, at maximum available resolution, resulting in 24,133 jpg images uploaded without mask or padding onto the Labelbox annotation platform (www.labelbox.com; last accessed

Table 3
Dataset description. Details of the number of annotations for each institution participating in the study: Cologne, Milan, Turin, Weill-Cornell, Mainz, Maastricht, Budapest.

Institution	Cortex	Medulla	Capsule/Other	Glomerulus	Artery	Arteriole
Cologne	2236	701	2363	10796	2124	14089
Milan	160	25	32	650	86	946
Turin	166	132	102	1040	258	1490
Weill-Cornell	363	62	294	2218	530	3113
Mainz	230	114	220	1087	182	1264
Maastricht	194	102	60	1509	239	2063
Budapest	128	62	52	700	168	912

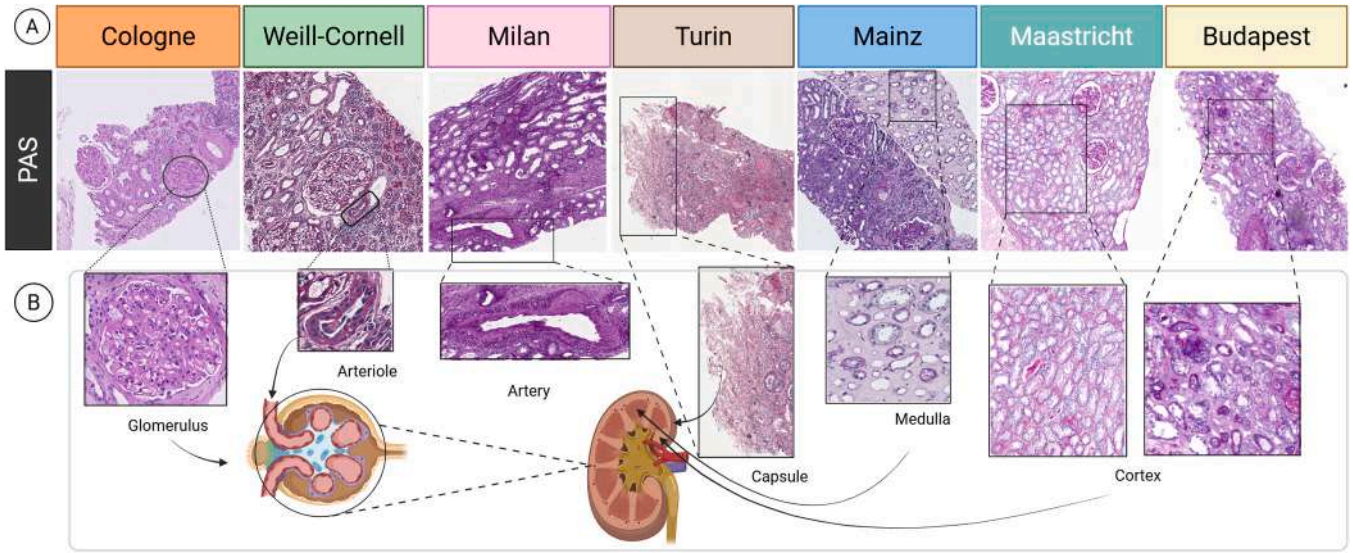

Fig. 1. Exemplary nephropathology images from the dataset. Images from the seven institutions participating in the study are represented in this figure: Cologne, Weill-Cornell, Milan, Turin, Mainz, Maastricht, Budapest. (A) PAS-stained images from the seven institutions. (B) Details from the images. Note that this study also included HE, trichrome and silver stainings.

September 9, 2025). JUB provided mutually exclusive expert labels (yes_tma/no/inconclusive) for all extracted crops.

These expert labels and the crop images of glomeruli, arteries, and arterioles, were then used for the TMA classification module described in Section 3.3.6.

3.3. Workflow

The workflow employed for the study is shown in Fig. 3.

First, an experienced nephropathologist (JUB) produced annotation masks in QuPath and instance-level TMA classification labels in Labelbox, as described in Sections 3.2.1 and 3.2.2.

To create the segmentation dataset, we realized a Groovy script that uses *LabeledImageServer* and *TileExporter*, setting different export resolutions according to the structures of interest. For the segmentation of the main anatomical compartments of the kidney, i.e., Cortex, Medulla, and Capsule/Other (that we refer to as "coarse" segmentation in the following), we employed an export resolution of 4 μm/pixel, while for the segmentation of the "fine" structures, i.e., Glomerulus, Artery, Arteriole, we employed an export resolution of 1 μm/pixel. The number of instances for each structure is reported in Section 3.2, with indication of how many samples were employed for the training, validation, and testing portions, and how they are scattered among the participating institutions.

We used this dataset to train and evaluate tile-level semantic segmentation models, as described in Sections 3.3.1 and 3.3.2. The results were then aggregated at WSI-level, and post-processed, as detailed in Sections 3.3.3 and 3.3.4.

In Section 3.3.5, we explained how the proposed workflow can be used for a Nephropathology Support System. Finally, in Section 3.3.6, we provide an application for TMA screening, by realizing an instance-level TMA classification module for the "fine" compartments considered.

3.3.1. Semantic segmentation models

We adopted the segmentation architectures from two different popular state-of-the-art DL segmentation frameworks: SegmentationModels (https://github.com/qubvel-org/segmentation_models.pytorch) and MMSegmentation (https://github.com/open-mmlab/mmsegmentation).

Deep convolutional neural networks are capable of learning hierarchical feature representations by integrating low- to high-level features across multiple layers. However, as network depth increases, optimization becomes more challenging, and degradation of training accuracy may occur—not due to overfitting, but rather due to the difficulty in propagating information through numerous nonlinear transformations. Deep Residual Networks (ResNets) mitigate this issue by introducing residual mappings with respect to the input of each layer (He et al., 2016).

To achieve robust segmentation performance, even with limited annotated data, (Ronneberger et al., 2015) proposed the U-Net architecture, which employs a symmetric encoder-decoder structure with skip connections that concatenate high-resolution feature maps from the encoder with upsampled outputs from the decoder. These skip connections facilitate precise localization. In our implementation, we adopted ResNet-34 as the encoder backbone within the U-Net framework, leveraging its residual connections to improve feature extraction and gradient flow during the training process.

In traditional convolutional architectures, the encoder progressively captures higher-level semantic features at the cost of spatial resolution. When only the deepest layer is used for object recognition, performance often degrades, particularly for small objects. To address this limitation and enable recognition across a range of scales, Feature Pyramid Networks (FPN) introduce a multi-scale architecture that leverages feature maps at different resolutions (Lin et al., 2017). FPN constructs a top-down pathway that progressively upsamples semantically rich, low-resolution features and combines them with higher-resolution, spatially precise features via lateral connections. This design enables the network to produce high-resolution feature maps with strong semantics at all scales, enhancing segmentation accuracy for variable object sizes.

To capture rich contextual information at multiple spatial scales, DeepLabV3+ integrates an Atrous Spatial Pyramid Pooling (ASPP) module within the encoder (Chen et al., 2018). This module employs parallel atrous convolution branches with varying dilation rates, enabling the network to effectively aggregate multi-scale contextual features. In our implementation, we utilize ResNet-34 as the encoder backbone with an output stride of 16, which balances spatial resolution with receptive field size. In the decoder, the encoder output is upsampled by a factor of 4 via bilinear interpolation and concatenated with corresponding low-level features from the encoder. This design facilitates precise object boundary refinement while preserving high-level semantic context in the final segmentation map.

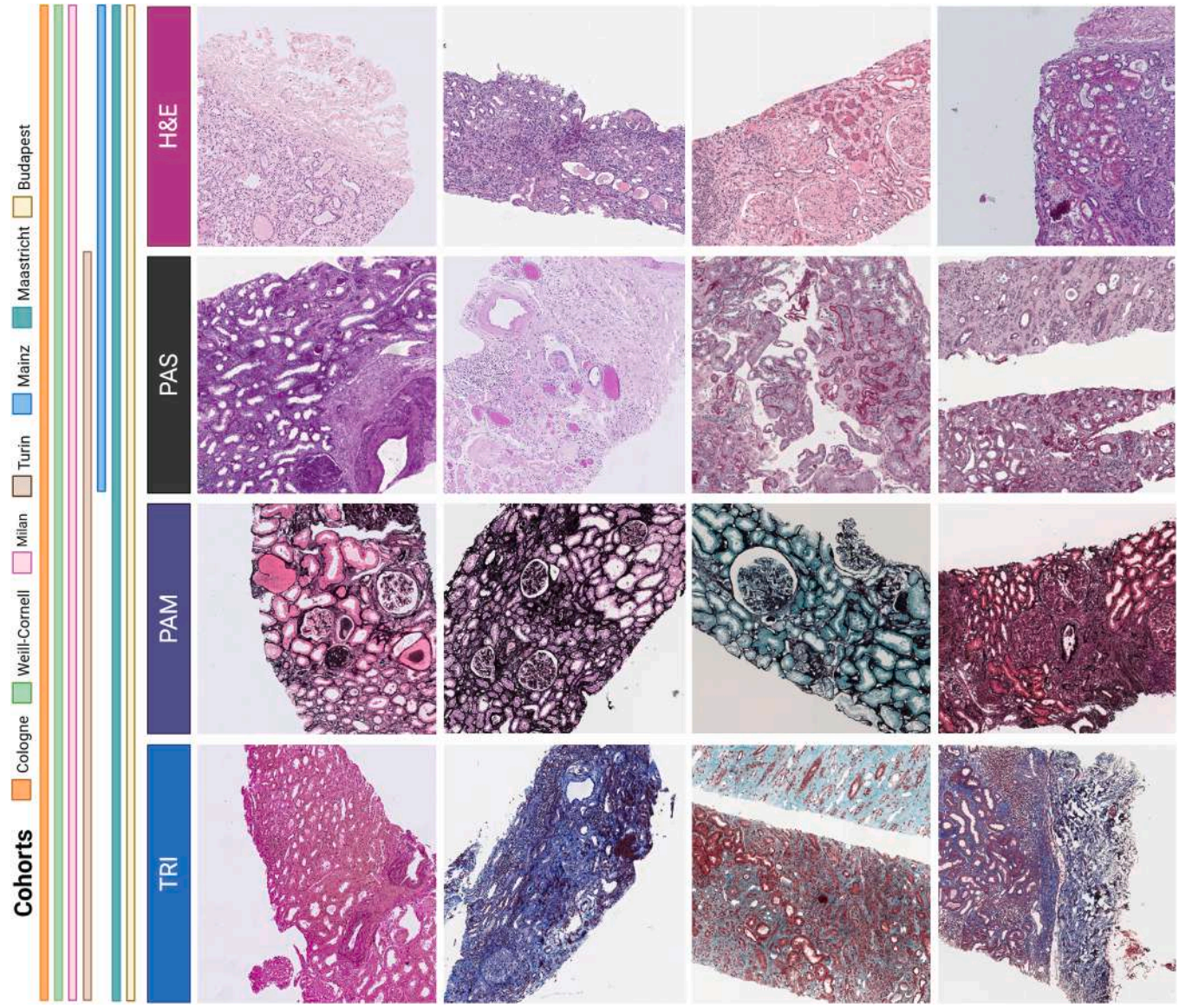

Fig. 2. Nephropathological stainings. The four main nephropathological stainings are herein represented: hematoxylin-eosin (HE), periodic-acid Schiff (PAS), periodic acid-silver methenamine (PAM), and trichrome (TRI) stainings. Note the large differences in tones of the same stainings from different institutions.

Masked-attention Mask Transformer (Mask2Former) is widely recognized as a versatile architecture designed to handle a broad range of segmentation tasks, including instance segmentation and panoptic segmentation (Cheng et al., 2022). In our implementation, we use a Swin Transformer (Liu et al., 2021) tiny (Swin-T) as the backbone, aiding in effective optimization convergence. The positional encoding for image patches is learnt during training using 128-dimensional features, rather than the traditional sinusoidal encoding. The pixel decoder incorporates Multiscale Deformable Attention layers that allow the model to focus on various parts of the image across different scales. The Transformer decoder utilizes Multi-Head Attention with 8 parallel attention heads, each learning different query points to improve the detection of small objects in the segmentation task.

We implemented the SegFormer architecture leveraging the Mix Transformer (MiT) as backbone, selecting the largest size variant (MiT-B5). SegFormer combines efficiently the long-range dependencies captured by transformers with local feature extraction capability of CNNs (Xie et al., 2021). The streamlined decoder consisted of multilayer perceptron (MLP) operators, which perform bilinear up-sampling and concatenation of features to predict the segmentation mask.

Unlike the architectures above, that leverage either pure transformer-based mechanisms or hybrid strategies, SegNeXt (Guo et al., 2022) is a convolutional network in which some computationally demanding tasks, such as heavy upsamplings, are simplified. The backbone was empowered with the Multi-Scale Context-Aware Attention (MSCAN) helping the model learn both fine-grained local information and contextual features. In our implementation, we leveraged MSCAN-T. The decoder was equipped with matrix decomposition layers which factorized the input representation and reduced the computational cost of self-attention by implementing the Hamburger (Geng et al., 2021) head.

We compared the following architectures: U-Net (ResNet-34 as backbone), FPN (ResNet-34 as backbone), DeepLabV3+ (ResNet-34 as backbone), Mask2Former (Swin-T as backbone), SegFormer (MiT-B5 as backbone), SegNeXt (MSCAN-T as backbone).

3.3.2. *Tile-level performance evaluation*

From a quantitative point of view, we evaluated the segmentation performance in terms of Recall, Precision, and Dice coefficient. The metrics have been calculated separately for the considered areas of

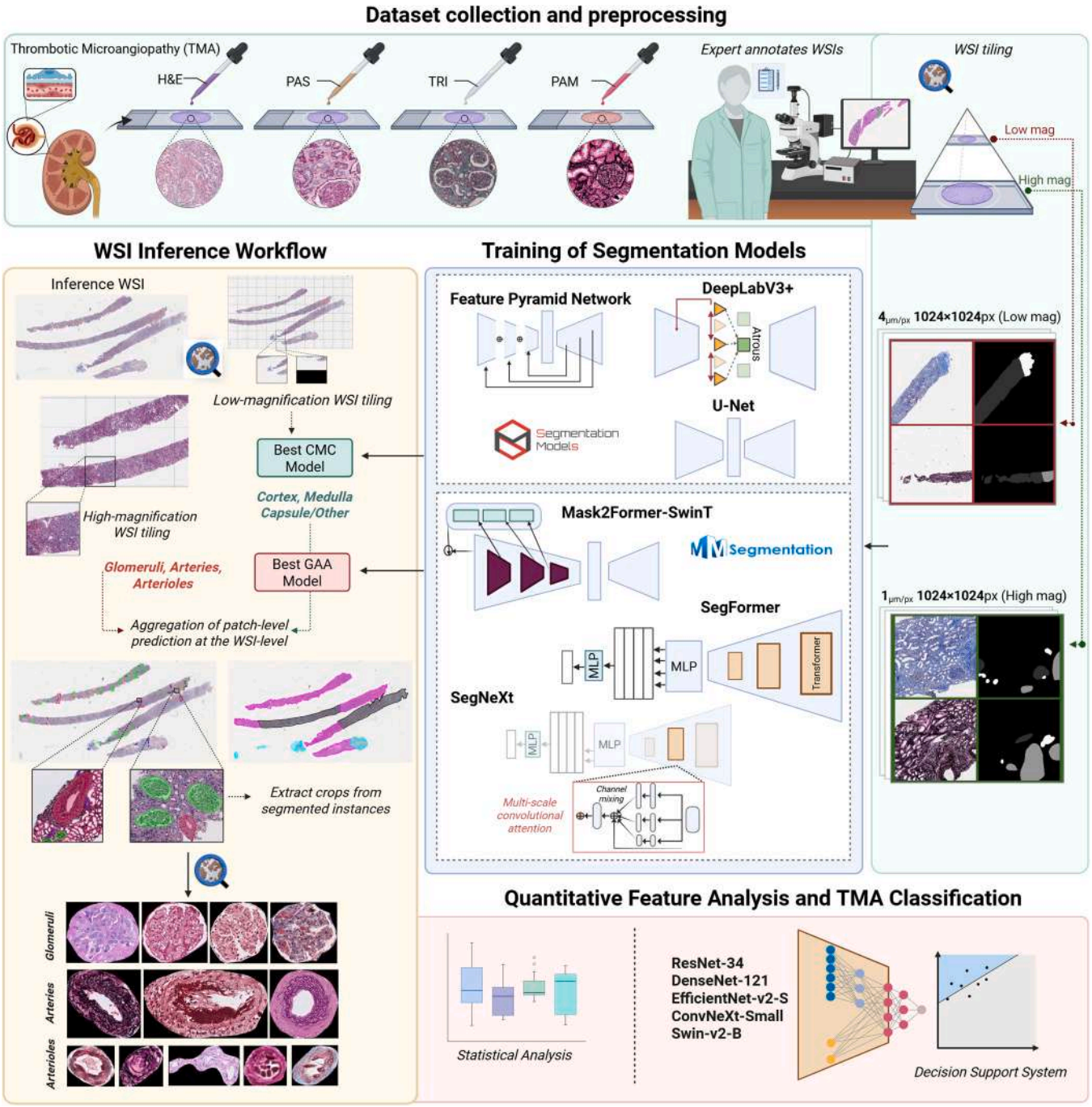

Fig. 3. Study Workflow. A cohort of biopsies with TMA or with other similar vasculopathies (Mimickers) resembling histological traits of TMA was collected. All four main nephropathological stainings were included (HE, PAS, PAM, TRI). An expert nephropathologist (JUB) annotated six areas of interest inside the WSIs: Cortex, Medulla, Capsule/Other, Glomerulus, Artery, Arteriole. A cascaded pipeline with semantic segmentation models working at two different magnitudes (4 μm/pixel for Cortex, Medulla, Capsule/Other and 1 μm/pixel for Glomerulus, Artery, Arteriole) was developed. Six different segmentation models were thoroughly compared. Results are then aggregated at WSI-level and displayed on QuPath through an integration between Python using the PAQUO library. The pipeline allows the export of segmented instances for revision by nephropathologist and/or further downstream analysis. Five models were adopted for classifying TMA at instance-level, for the Glomerulus and Artery/Arteriole compartments, providing nephropathological clinical screening support.

interest: Cortex, Medulla, Capsule/Other, Glomerulus, Artery, Arteriole.

Furthermore, we qualitatively compared the segmentation performance by observing the differences between the segmentation masks produced by the six considered models and with respect to the ground truth.

The best models at tile-level were selected for the end-to-end experiments at WSI-level.

3.3.3. WSI-level projection

This section describes the aggregation rule adopted to reconstruct WSI-level segmentation masks starting from tile-level predictions having overlapping regions. For each class (excluding the background), the WSI-level mask is generated using a weighting strategy that helps resolve ambiguities caused by overlapped tile predictions. The weighting function used, $W(x, y)$, is analogous to a 2D Heaviside Pi function, as described by the following equation:

$$W(x,y) = \begin{cases} 0.5, \text{ if } \left(\frac{x}{X} < \frac{1}{8}\right) \cup \left(\frac{y}{Y} < \frac{1}{8}\right) \cup \left(\frac{x}{X} > \frac{7}{8}\right) \cup \left(\frac{y}{Y} > \frac{7}{8}\right) \\ 1.0, \text{ if } \left(\frac{1}{8} \leq \frac{x}{X} \leq \frac{7}{8}\right) \cap \left(\frac{1}{8} \leq \frac{y}{Y} \leq \frac{7}{8}\right) \end{cases},$$

where $x \in [0, X-1], y \in [0, Y-1]$, are the spatial coordinates in the tile, and X, Y are the tile dimensions.

As auxiliary masks, a Count Mask and a Sum Mask are created for the WSI. The Count Mask, for each coordinate x,y, contains the number of tiles that include that coordinate. The Sum Mask, for each coordinate x, y, contains the weighted sum of the binary class masks from the tiles that include that coordinate, weighted using the kernel $W(x,y)$. Finally, WSI-level Average Mask is created by performing a pixelwise division between Sum Mask and Count Mask. Only pixels of the Average Mask having values greater than a threshold (we used 0.15 in our experiments) are retained as prediction for that class. Note that at this stage, binary class masks are employed for each class, and hence a pixel can belong to more than one class. This issue is addressed in the post-processing.

3.3.4. Post-processing

The post-processing phase was aimed at handling the aggregation of binary masks, filtering and spatial alignments of different resolution compartments. Initially, the lower-resolution tissue masks, i.e., of "coarse" compartments were upscaled to match the resolution of the "fine" compartments, i.e., Glomerulus, Artery, Arteriole.

Artery and Arteriole masks were only retained if belonging to valid tissue regions (i.e., are located inside Cortex, Medulla, or Capsule/Other). An area-based filtering, with a threshold depending on the compartment, was used to remove erroneously segmented objects.

Subsequently, a hierarchical label suppression strategy using nephropathology principles was applied to resolve overlapping regions. At the "coarse" level, Cortex regions were prioritized over Medulla and Capsule/Other areas. For the "fine" compartments, Glomerulus class was prioritized over Artery and Arteriole classes.

A watershed-based algorithm was developed to decide whether a Glomerulus mask may correspond to multiple adjacent (or "kissing") glomeruli that have to be separated into different connected objects. First, connected components were identified in the binary glomerulus mask using morphological labeling. For each component, a region of interest (ROI) was extracted based on its bounding box. The Euclidean distance transform was computed within the ROI to highlight the centers of individual structures. Using this distance map, local maxima were detected as candidate glomerular centers, constrained by a minimum distance threshold (40 pixels at 1 μm/pixel) and a maximum number of peaks (we used 3, as it is unlikely that more than 3 glomeruli are touching at the same time). These peak coordinates were used as seeds for segmenting the area of influence starting from the inverted distance map. If only one object was detected (i.e., no internal peaks found), the original ROI was retained. In cases where multiple watershed segments were identified, each label was post-processed using morphological erosion to dissect the touching glomeruli, before being reintegrated into the global WSI mask as individual glomeruli.

Finally, we provide Python and Groovy interfaces for exporting objects of interest. Furthermore, we made available a function for exporting also the tiles composing "coarse" regions such as the Cortex, which can be used for further analyses.

3.3.5. Nephropathology support system

We developed our segmentation pipeline as a support system for nephropathologists, either in their diagnostic practice, where it could be used to highlight or count specific tissue segments or for downstream analysis with expert labels or as an input for machine learning classifiers or regression models. To this end, the weights of the trained model are freely available on HuggingFace: https://huggingface.co/NePathTK. The entire **NePathTK** pipeline is publicly available on GitHub: https://github.com/Nicolik/NePathTK. Finally, a QuPath integration was realized by exploiting the PAQUO (PAthological QUpath Obsession) (https://paquo.readthedocs.io/en/latest/) library for visualization and further exploitation of the segmentation results with an application commonly used by pathologists. To export the crops for diagnostic examinations, both Python and Groovy interfaces were made available. Example features that can be extracted from such crops are described in Supplementary Methods. The next subsection describes the classification module for TMA at instance-level for such crops.

3.3.6. TMA classification

We employed classification models from the state-of-the-art DL package torchvision of PyTorch. We trained five models for the TMA instance-level classification: ResNet-34, DenseNet-121, EfficientNet-v2-S, ConvNeXt-Small, Swin-v2-B. For each of these models, two separate configurations were considered: one using only the glomerular and the other using artery plus arteriole compartments. This was done because from a nephropathological viewpoint, the manifestations of TMA are quite similar in arteries and arterioles; apart from the eponymous thrombi, they are very different in arteries/arterioles compared to glomeruli. We used masked compartments in order to avoid the leak of information between the compartments, e.g. in an image showing a glomerulus without evidence of TMA but with a microthrombus in an afferent arteriole.

ResNet-34 (He et al., 2016) is introduced in Section 3.3.1, since it is the backbone of the CNN-based semantic segmentation models we considered.

DenseNet (Huang et al., 2017) is a fully convolutional architecture where each layer receives as input the concatenation of all preceding feature maps, while its own output is propagated to all subsequent layers. Consequently, instead of the conventional L connections for L layers, DenseNet establishes $L(L+1)/2$ direct connections in a feed-forward manner. This dense connectivity pattern facilitates efficient feature reuse, mitigates the vanishing gradient problem by ensuring uninterrupted information and gradient flow, and significantly reduces the number of parameters by avoiding redundant feature relearning. We adopted DenseNet-121 in our experiments.

EfficientNet-v2 (Tan and Le, 2021) is a fully convolutional architecture optimized for both training speed and parameter efficiency, in contrast to DenseNet or the original EfficientNet (Tan and Le, 2019), which required a larger number of parameters. Its design leverages a progressive learning strategy, wherein the model is initially trained on lower-resolution images with weaker regularization during the early epochs. As training progresses, the input resolution is gradually increased, and stronger regularization techniques are introduced. This approach accelerates convergence and reduces the computational cost, without sacrificing predictive accuracy. We exploited EfficientNet-v2-S in our experiments.

ConvNeXt (Liu et al., 2022) is a convolutional architecture that integrates design paradigms from hierarchical vision Transformers (e.g., Swin) into a ResNet-style backbone, while retaining purely convolutional operations without attention modules. Each ConvNeXt block adopts an inverted bottleneck configuration, where the hidden dimension of the feed-forward MLP is expanded to four times the input dimension, followed by depthwise convolution and layer normalization. This configuration substantially enhances representational capacity relative to the modest increase in FLOPs, making the architecture both efficient and expressive. ConvNeXt does not require specialized modules such as shifted window attention or relative position biases as in the hierarchical ViT. We utilized ConvNeXt-Small in our experiments.

The Swin Transformer architecture has demonstrated strong performance as a backbone for fine-grained vision tasks. In the context of distinguishing subtle phenotypic differences between TMA and mimicker instances, higher-level semantic decomposition plays a critical role. The Swin-v2 (Liu et al., 2022) architecture introduces

Table 4
Quantitative results for Cortex, Medulla, Capsule/Other on the Validation Set. Metrics: Recall, Precision, Dice.

Model	Cortex			Medulla			Capsule/Other		
	Recall	Precision	Dice	Recall	Precision	Dice	Recall	Precision	Dice
Mask2Former	0.906	**0.936**	0.921	0.793	0.829	**0.810**	0.865	0.857	**0.861**
SegNeXt	0.903	0.931	0.917	**0.806**	0.770	0.788	0.841	**0.867**	0.854
SegFormer	**0.924**	0.919	**0.922**	0.729	**0.871**	0.794	0.849	0.861	0.855
U-Net	0.899	0.927	0.913	0.694	0.836	0.759	**0.903**	0.640	0.749
FPN	0.913	0.923	0.918	0.750	0.807	0.777	0.851	0.848	0.849
DeepLabV3+	0.912	0.933	**0.922**	0.794	0.795	0.794	0.848	**0.867**	0.858

Table 5
Quantitative results for Cortex, Medulla, Capsule/Other on the Test Set. Metrics: Recall, Precision, Dice.

Model	Cortex			Medulla			Capsule/Other		
	Recall	Precision	Dice	Recall	Precision	Dice	Recall	Precision	Dice
Mask2Former	0.905	0.893	0.899	0.769	0.900	0.829	0.834	0.800	**0.816**
SegNeXt	0.916	0.890	**0.903**	**0.772**	0.908	**0.834**	0.785	**0.845**	0.814
SegFormer	**0.925**	0.878	0.901	0.743	0.939	0.830	0.787	0.818	0.803
U-Net	0.906	0.856	0.880	0.574	**0.944**	0.714	**0.862**	0.416	0.561
FPN	0.915	0.849	0.880	0.651	0.937	0.768	0.767	0.750	0.759
DeepLabV3+	0.864	**0.896**	0.880	0.745	0.900	0.815	0.715	0.825	0.766

modifications over the original Swin Transformer (Liu et al., 2021). Specifically, the normalization layer is repositioned before the residual blocks, a configuration termed *res-post-norm*, which improves training stability in deeper networks. In addition, Swin-v2 employs scaled cosine attention, mitigating the risk of extremal values in the dot product between query and key feature vectors and thereby enhancing numerical stability during attention computation. In our experiments, we leveraged on Swin-v2-B.

The models were evaluated in terms of Accuracy, Precision, Recall, and Specificity. Receiver Operating Characteristic (ROC) and Precision Recall (PR) curves, with the associated Area Under ROC curve (AUROC) and Area Under PR curve (AUPRC), were also used to assess the TMA instance-level classification models.

3.4. Ethical permissions

The retrospective use of de-identified data for epidemiological research projects like this in Cologne is permitted by state law in North Rhine-Westphalia (Berufsordnung der nordrheinischen Ärztinnen und Ärzte), Germany (https://www.aekno.de/fileadmin/user_upload/aekno/downloads/2025/berufsordnung-2025.pdf; last accessed October 2025).

4. Results

4.1. "Coarse" segmentation: cortex, medulla, capsule/other

For the validation set, quantitative results for Cortex, Medulla, Capsule/Other are reported in Table 4, including the metrics of Recall, Precision and Dice coefficient. For Cortex Segmentation, SegFormer and DeepLabV3+ exhibit a Dice coefficient of 0.922. For Medulla Segmentation, Mask2Former emerges with a Dice coefficient of 0.810. For Capsule/Other, Mask2Former achieved a Dice coefficient of 0.861.

For the test set, quantitative results for Cortex, Medulla, Capsule/Other are reported in Table 5 in terms of Recall, Precision and Dice coefficient. For Cortex Segmentation, SegNeXt showed a Dice coefficient of 0.903. For Medulla Segmentation, SegNeXt displayed a Dice coefficient of 0.834. For Capsule/Other, Mask2Former had a Dice coefficient of 0.816.

A qualitative comparison featuring the six considered semantic segmentation models is shown in Fig. 4 for images of the validation set, and in Fig. 5 for images of the test set.

4.2. "Fine" segmentation: glomerulus, artery, arteriole

For the validation set, quantitative results for Glomerulus, Artery, Arteriole are reported in Table 6 with the metrics of Recall, Precision and Dice coefficient. For Glomerulus segmentation, Mask2Former had a Dice coefficient of 0.936. For Artery segmentation, Mask2Former had a Dice coefficient of 0.827. For Arteriole segmentation, Mask2Former had a Dice coefficient of 0.673.

For the test set, quantitative results for Glomerulus, Artery, Arteriole are reported in Table 7 with the same metrics of Recall, Precision, Dice coefficient. For Glomerulus segmentation, Mask2Former had a Dice coefficient of 0.922. For Artery segmentation, SegFormer achieved a Dice coefficient of 0.822. For Arteriole segmentation, Mask2Former had a Dice coefficient of 0.553.

A qualitative comparison featuring the six semantic segmentation models is shown in Fig. 6 for images of the validation set, and in Fig. 7 for images of the test set.

4.3. Nephropathology decision support system

The integration of our pipeline with the popular software QuPath for histopathology data visualization and analysis is shown in Fig. 8. Mask2Former displayed robust segmentation performances in all compartments, attaining the best Dice coefficients for Glomerulus, Artery, Arteriole, Medulla, and Capsule/Other on the Validation Set, and for Glomerulus, Arteriole, and Capsule/Other, on the Test Set, as described in see Sections 4.1 and 4.2. Accordingly, we recommend Mask2Former as the default segmentation model for both "coarse" and "fine" compartments in our pipeline, while keeping all developed models available for use.

Features extracted from segmented compartments can be used to perform statistical analyses (for instance, based on TMA vs Mimicker condition), as reported in Figure S1 and Figure S2 in Supplementary Results. A comparison between the features extracted by expert- and Mask2Former-segmented instances for the Budapest cohort is reported in Figure S3.

4.4. TMA classification

Fig. 9 and Fig. 10 portray the ROC and PR curves of the five evaluated deep learning models (ResNet-34, DenseNet-121, EfficientNet-v2-S, ConvNeXt-Small, and Swin-v2-B) trained on the Glomerulus and Artery

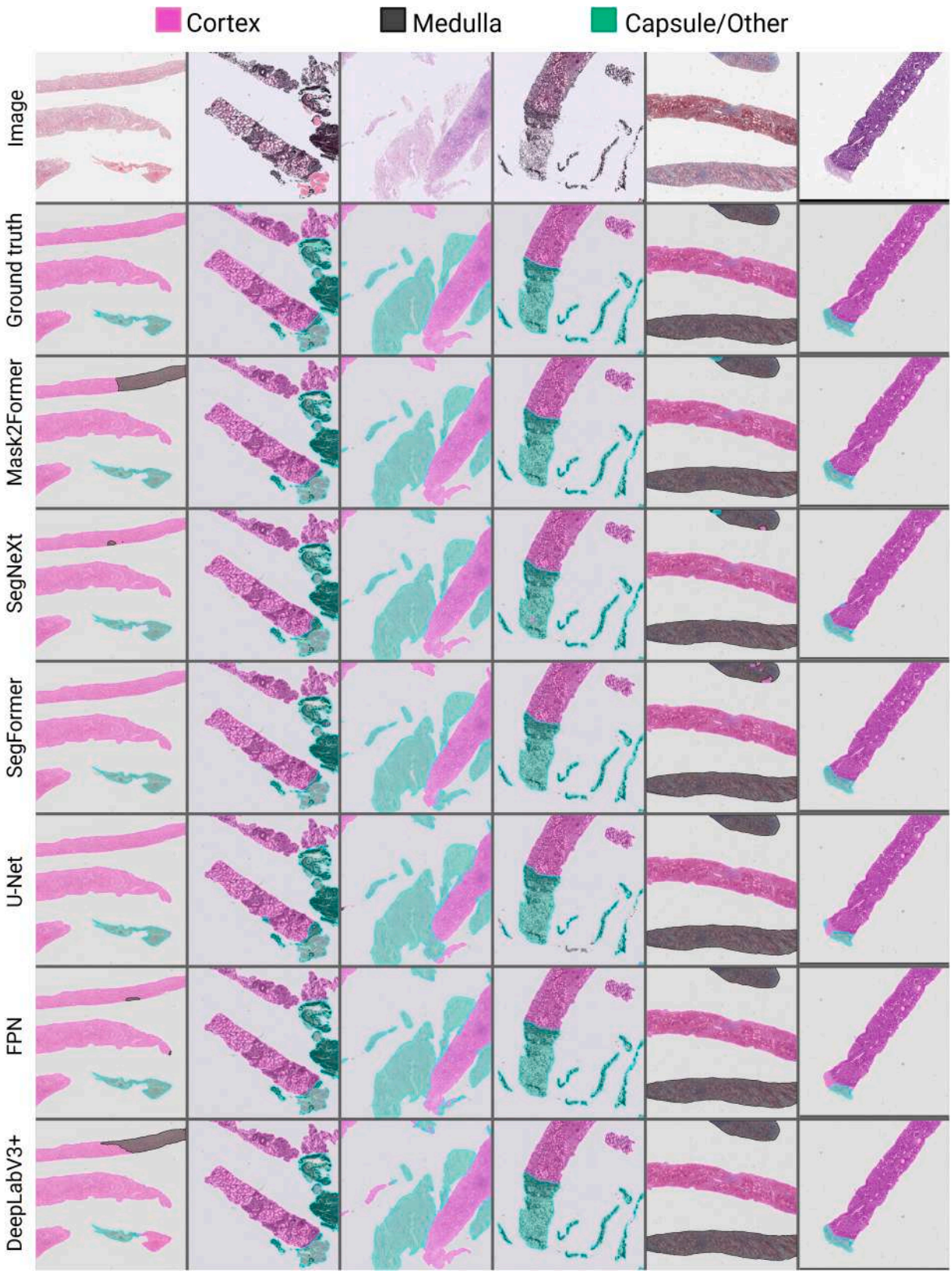

Fig. 4. Comparison of qualitative results for Cortex, Medulla, Capsule/Other on the Validation Set. The top row shows the original images, the second row the ground truth masks, and the other lines show the masks obtained using the different Deep Learning models.

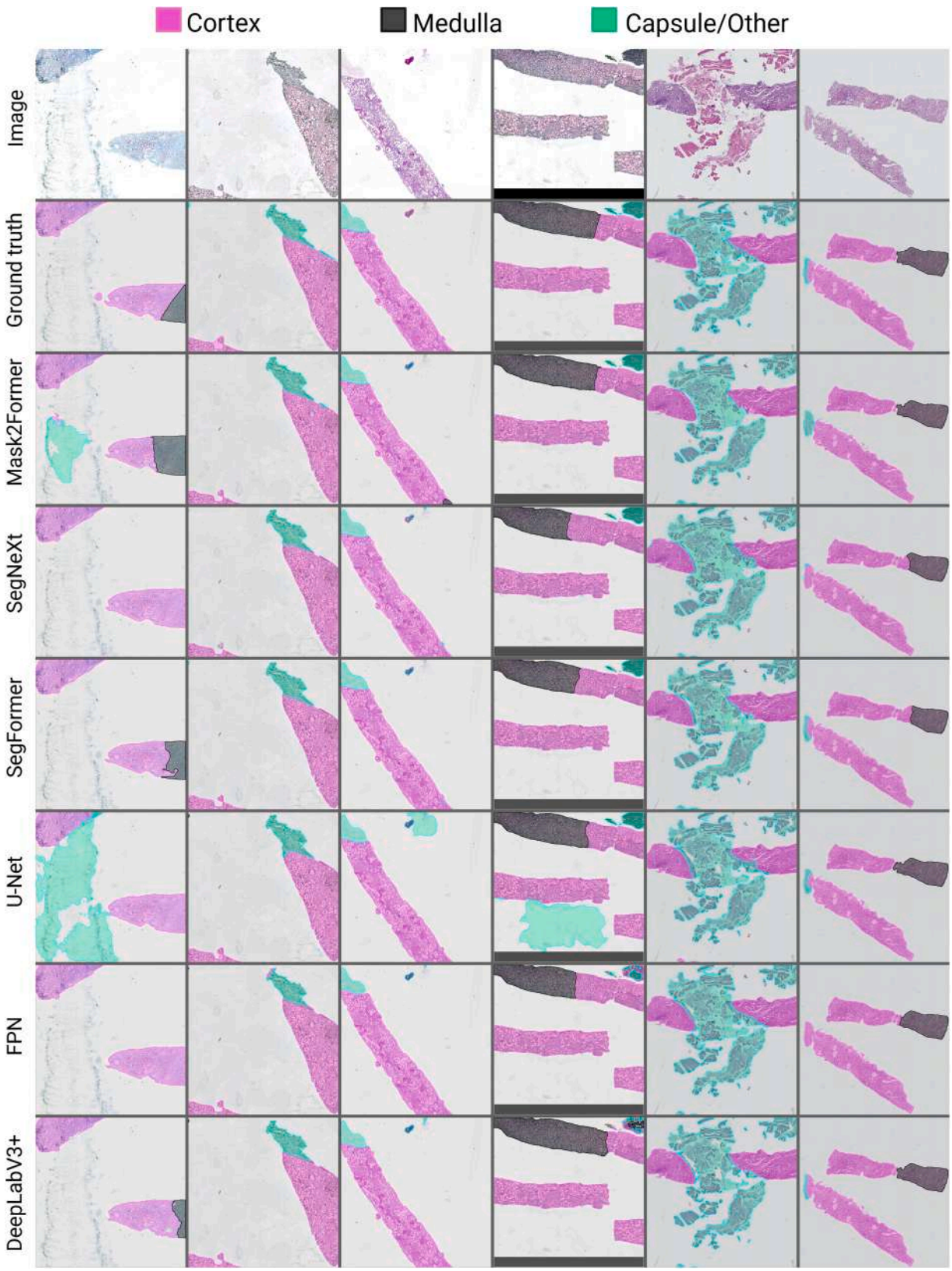

Fig. 5. Comparison of qualitative results for Cortex, Medulla, Capsule/Other on the Test Set. The top row shows the original images, the second row the ground truth masks, and the other lines show the masks obtained using the different Deep Learning models.

Table 6
Quantitative results for Glomerulus, Artery, Arteriole on the Validation Set. Metrics: Recall, Precision, Dice.

Model	Glomerulus			Artery			Arteriole		
	Recall	Precision	Dice	Recall	Precision	Dice	Recall	Precision	Dice
Mask2Former	**0.931**	0.942	**0.936**	**0.810**	0.845	**0.827**	**0.630**	0.722	**0.673**
SegNeXt	0.916	0.945	0.930	0.775	0.857	0.814	0.513	0.763	0.614
SegFormer	0.914	0.944	0.929	0.792	**0.861**	0.825	0.530	0.757	0.624
U-Net	0.887	**0.960**	0.922	0.787	0.734	0.760	0.471	0.789	0.590
FPN	0.892	0.955	0.923	0.737	0.825	0.779	0.506	0.757	0.607
DeepLabV3+	0.903	0.949	0.926	0.742	0.805	0.772	0.446	**0.794**	0.571

Table 7
Quantitative results for Glomerulus, Artery, Arteriole on the Test Set. Metrics: Recall, Precision, Dice.

Model	Glomerulus			Artery			Arteriole		
	Recall	Precision	Dice	Recall	Precision	Dice	Recall	Precision	Dice
Mask2Former	**0.906**	0.939	**0.922**	0.798	0.788	0.793	**0.507**	0.608	**0.553**
SegNeXt	0.886	0.955	0.919	0.796	**0.844**	0.819	0.396	0.719	0.511
SegFormer	0.886	0.951	0.917	**0.813**	0.830	**0.822**	0.409	0.721	0.522
U-Net	0.816	**0.963**	0.884	0.652	0.839	0.734	0.392	0.686	0.499
FPN	0.847	0.958	0.899	0.733	0.821	0.775	0.397	0.724	0.513
DeepLabV3+	0.862	0.955	0.906	0.753	0.807	0.779	0.334	**0.765**	0.465

plus Arteriole compartments. Table 8 and Table 9 detail the corresponding quantitative metrics (accuracy, precision, recall, and specificity) on both validation and test subsets.

For the Glomerulus compartment, the highest validation accuracy was achieved by DenseNet-121 with a value of 0.791, while ConvNeXt-Small obtained the best accuracy on the test set with 0.724. The best validation AUROC was reached by EfficientNet-v2-S with 0.811, and ConvNeXt-Small again achieved the highest AUROC on the test set with 0.778. For AUPRC, Swin-v2-B performed best on validation with 0.603, whereas ConvNeXt-Small reached the highest value on the test set with 0.700.

For the Artery plus Arteriole compartments, Swin-v2-B achieved the best validation accuracy with 0.801, also the top test accuracy with 0.841. EfficientNet-v2-S yielded the best validation AUROC with 0.787, while Swin-v2-B obtained the highest test AUROC with 0.796. Regarding AUPRC, EfficientNet-v2-S performed best on validation with 0.610, and Swin-v2-B achieved the top score on the test set with 0.523.

5. Discussion

Nephropathologists integrate findings from different tissue compartments on the typical set of four main nephropathological stainings, i.e., TRI, PAS, PAM and HE, which all provide unique morphological information about tissue structure and composition. Morphological assessment of kidney biopsy WSIs is deemed crucial for differential diagnosis, prognosis and treatment selection especially in rare diseases like TMA, where diagnostic findings can be made in arteries, arterioles, and glomeruli, with secondary findings in the other compartments of cortex tubulointerstitium and medulla. Although clinical decision-making deserves thorough WSIs inspection and information merging from multiple stainings, level sections and under integration of clinical findings, an adequate DL-based quantitative framework is still lacking. The proposed automated image analysis framework enables to reliably isolate compartments from different nephropathological stainings and demonstrates considerable robustness to inter-departmental staining variations. Such compartments crops can then be used for feature engineering or other downstream tasks like classification, either individually or within multiple instance learning frameworks. Previous studies have typically focused on a single tissue compartment (Bueno et al., 2020), a single imaging modality (Bouteldja et al., 2021), or both (Altini et al., 2020b,a), limiting their applicability for comprehensive, first-pass assessment of kidney biopsies during early diagnostic workflows. Our segmentation pipeline, usable on all relevant stains, allows to process several WSIs for each case. Our models were trained and validated both on near-normal and severely altered renal compartments typical for TMA and the broad range of vasculopathies and vasculitis. Crucially, to the best of our knowledge, ours is only the second such segmentation model differentiating between arteries and arterioles. This is of importance because these segments are quite distinct in their pathobiology and pathomorphology, as is also reflected in the Banff Classification of renal transplant pathology, where it is of critical importance for the diagnostic process (Roufosse and Simmonds, 2018); since (Lutnick et al., 2022) only give the Matthew's Correlation Coefficient for the hold-out segmentation performance of their model on arteries (0.86) and arterioles (0.66), no direct comparison with our model is possible. Our model has the advantage of a larger training set of arterial and arteriolar annotation masks from more institutions.

Different DL architectures have been used to build a benchmark for the segmentation task considered. CNN-based U-Net, FPN, and DeepLabV3+ showed reliable baseline performance across all compartments, while transformer-based models like SegFormer and Mask2Former achieved superior Dice scores, especially in delineating complex vascular structures. Mask2Former emerged as the model having greater consistency across institutions and that was better suited to recognize the "fine" structures. Our findings confirm that self-attention-based Transformers perform better than traditional CNN models.

The new and rather large training set of high-quality expert annotations could be capable of delivering superior results over a wide range of stainings and diseases including near-normal biopsies both from native and transplant kidneys. We invite the nephropathology and nephrology communities to compare our segmentation pipeline to others on their datasets of interest.

Developed for the application on large WSI repositories for TMAs and other vasculopathies, a robust and accurate segmentation pipeline like ours can be used for generating tissue compartment crops for a range of downstream applications. These include the training of foundation models on unlabeled datasets, weakly and fully supervised learning, on WSI- or biopsy-level. Our integration of NePathTK into the commonly used QuPath software allows visualization and review of the segmentation results even for nephropathologists without coding expertise. As such, it could be integrated as a clinical support tool into nephropathology workflows for highlighting and counting e.g. glomeruli or arterioles. We expect NePathTK to play an important role in personalized medicine for rare nephropathies (Cazzaniga et al., 2023).

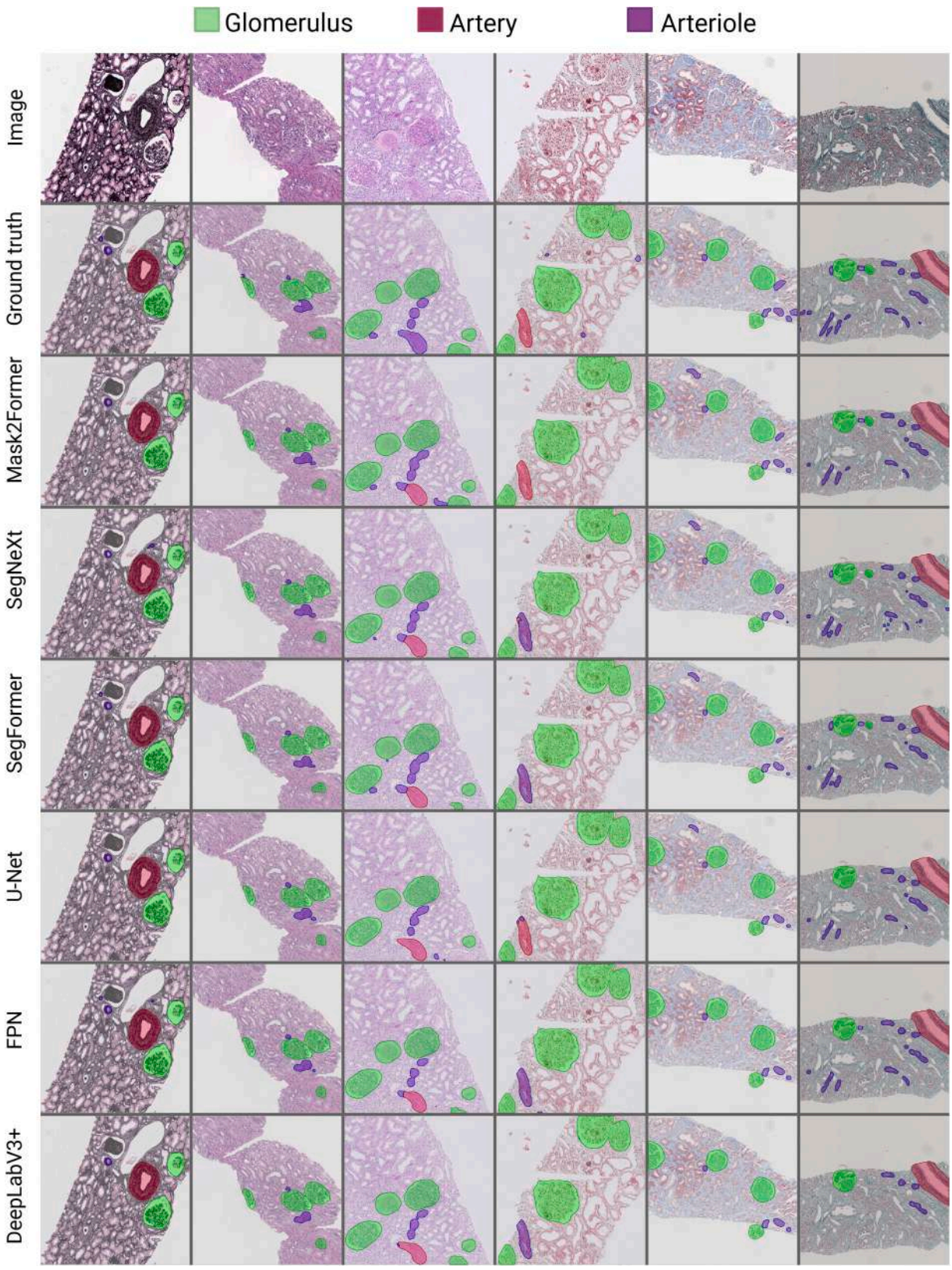

Fig. 6. Comparison of qualitative results for Glomerulus, Artery, Arteriole on the Validation Set. The top row shows the original images, the second row the ground truth masks, and the other lines show the masks obtained using the different Deep Learning models.

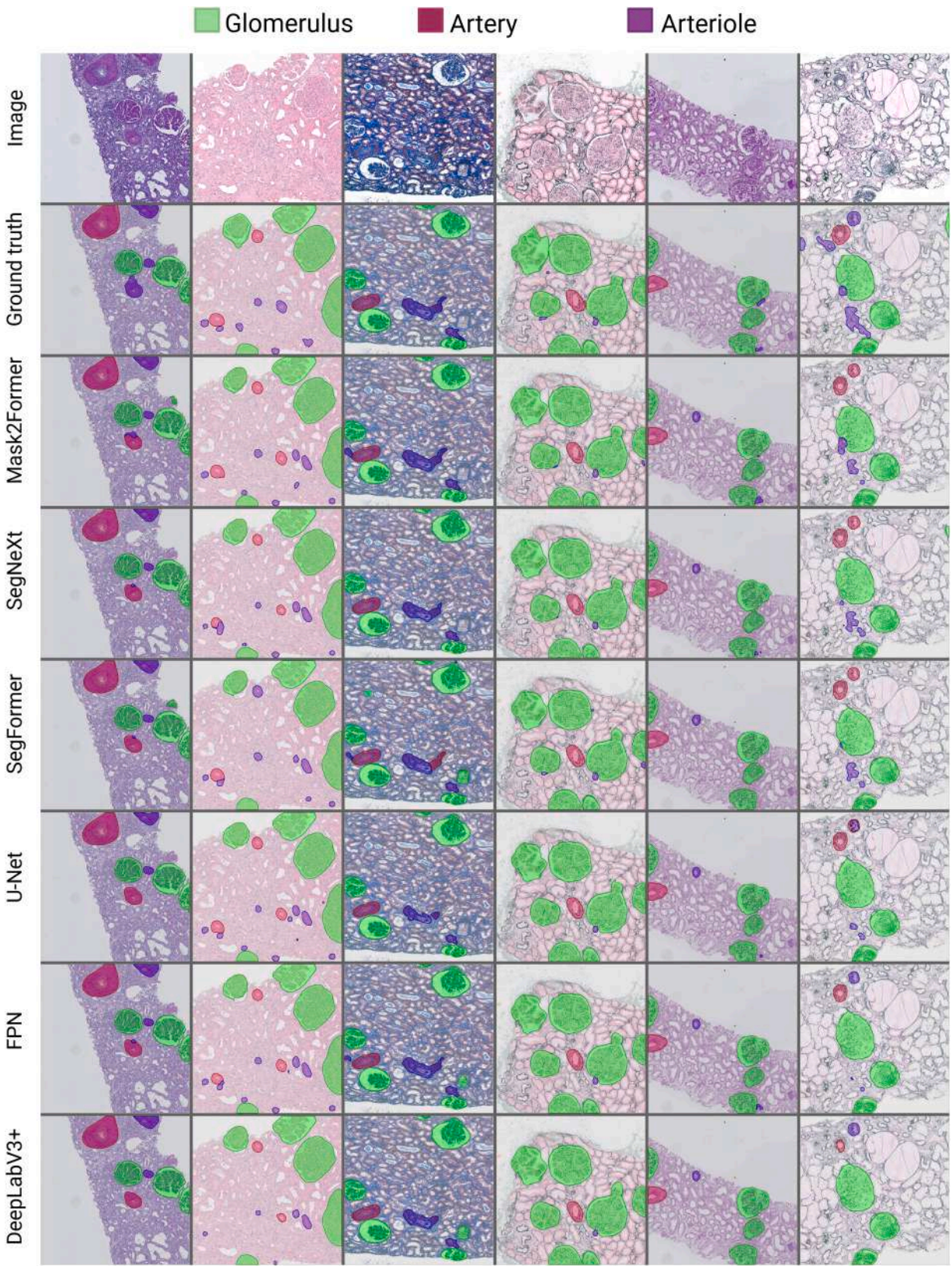

Fig. 7. Comparison of qualitative results for Glomerulus, Artery, Arteriole on the Test Set. The top row shows the original images, the second row the ground truth masks, and the other lines show the masks obtained using the different Deep Learning models.

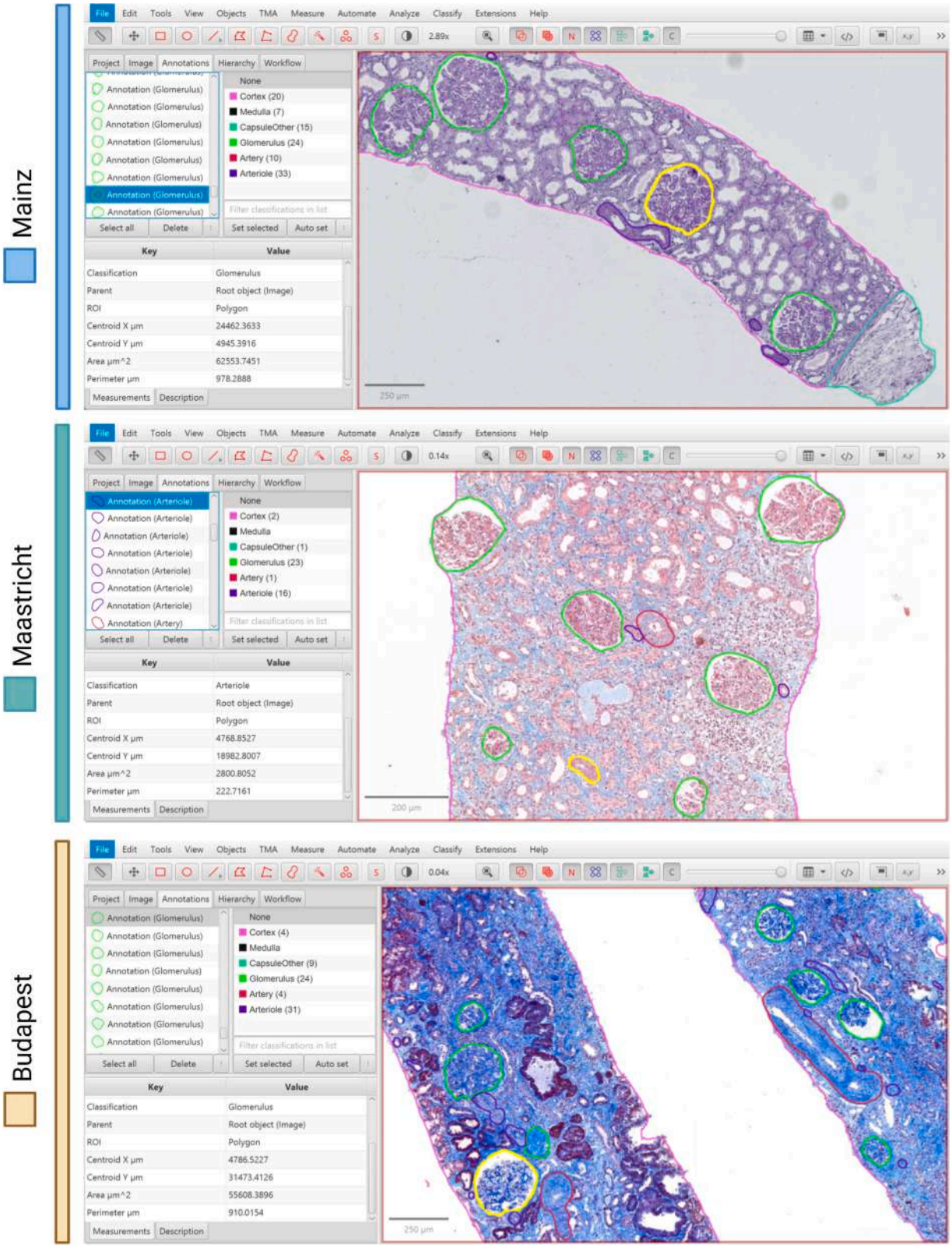

Fig. 8. QuPath integration. Results on the test cohorts. Top: Mainz; Middle: Maastricht; Bottom: Budapest.

As an example of such a downstream application, we provide a screening tool that could be clinically useful for nephropathologists. The comparative analysis of TMA instance-level classification models highlights the heterogeneous performance across architectures and tissue compartments.

In the Glomerulus compartment, DenseNet-121 showed best validation accuracy, but ConvNeXt-Small demonstrated more robust performance on the independent test set, achieving the highest Accuracy, AUROC, and AUPRC. This suggests that convolutional models incorporating Transformer-inspired design principles may better capture subtle

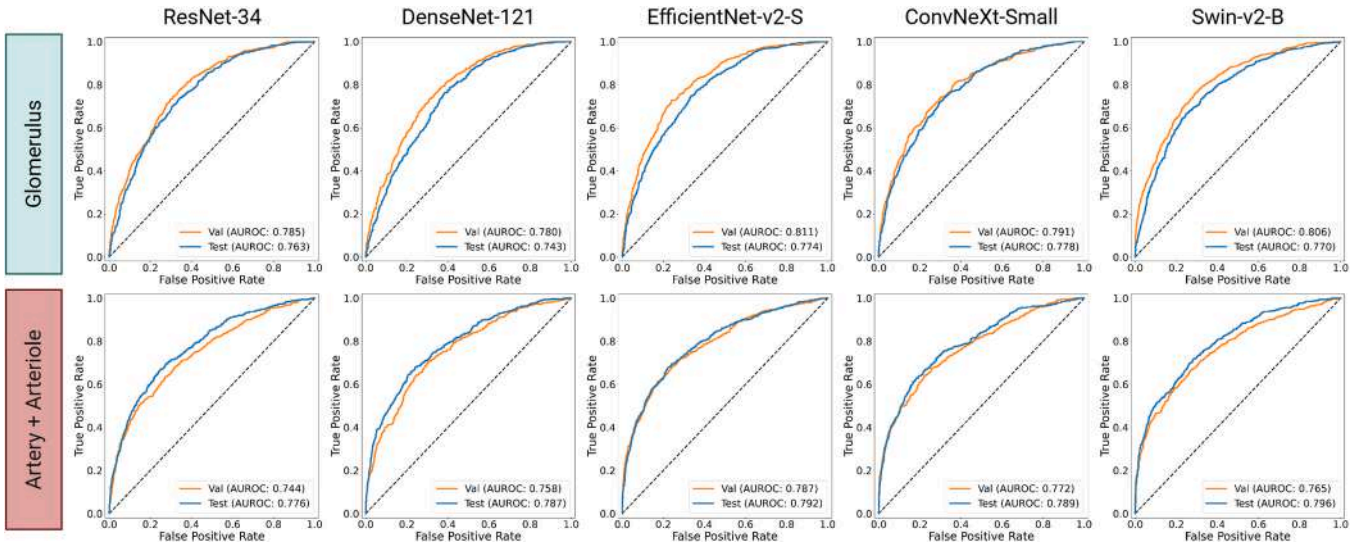

Fig. 9. ROC curves of five different models for instance-level TMA Classification. Top row: Glomerulus compartment; Bottom row: Artery plus Arteriole compartment.

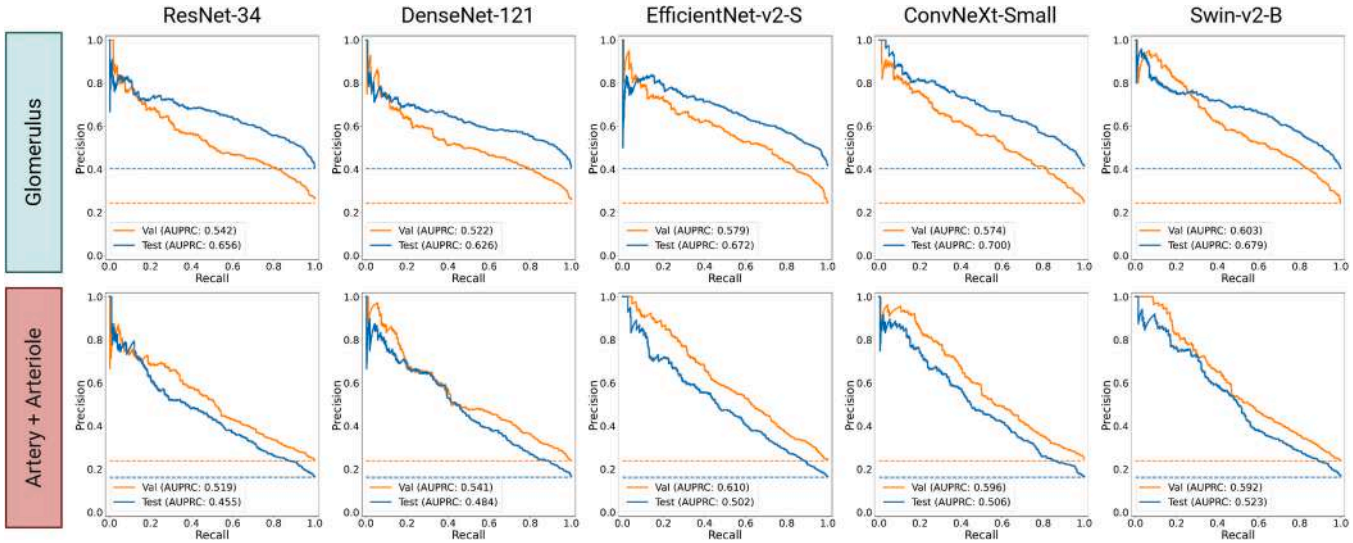

Fig. 10. PR curves of five different models for instance-level TMA Classification. Top row: Glomerulus compartment; Bottom row: Artery plus Arteriole compartment.

Table 8
Classification Metrics for models trained on Glomerulus compartment. Results are reported in terms of accuracy, precision, recall, specificity, for both Validation (Val) and Test subsets.

Model	Subset	Accuracy	Precision	Recall	Specificity
ResNet−34	Val	0.731	0.685	0.464	0.881
	Test	0.716	0.646	0.647	**0.762**
DenseNet−121	Val	**0.791**	0.332	**0.634**	0.814
	Test	0.671	0.336	0.687	0.667
EfficientNet-v2-S	Val	0.757	**0.713**	0.500	**0.893**
	Test	0.700	**0.657**	0.620	0.759
ConvNeXt-Small	Val	0.788	0.548	0.565	0.856
	Test	**0.724**	0.554	0.698	0.736
Swin-v2-B	Val	0.788	0.467	0.579	0.839
	Test	0.702	0.430	**0.718**	0.698

Table 9
Classification Metrics for models trained on Artery plus Arteriole compartment. Results are reported in terms of accuracy, precision, recall, specificity, for both Validation (Val) and Test subsets.

Model	Subset	Accuracy	Precision	Recall	Specificity
ResNet−34	Val	0.737	0.618	0.462	0.866
	Test	0.762	0.625	0.366	0.915
DenseNet−121	Val	0.777	0.490	0.536	0.844
	Test	0.837	0.478	0.501	0.899
EfficientNet-v2-S	Val	0.717	**0.743**	0.444	**0.898**
	Test	0.711	**0.733**	0.327	**0.931**
ConvNeXt-Small	Val	0.792	0.503	0.574	0.850
	Test	0.839	0.476	0.505	0.899
Swin-v2-B	Val	**0.801**	0.470	**0.609**	0.845
	Test	**0.841**	0.503	**0.512**	0.903

morphological differences in glomerular structures.

In the Artery/Arteriole compartments, Swin-v2-B consistently achieved the highest test accuracy, AUROC, and AUPRC, underscoring the utility of hierarchical attention-based architectures in modeling complex vascular distortions typical of TMA. EfficientNet-v2-S showed competitive results on the validation set, obtaining the best AUROC and

AUPRC, though these gains were not confirmed on the external test set.

Overall, these findings suggest that while convolutional backbones remain competitive, hybrid and Transformer-based models, such as ConvNeXt and Swin-v2-B, exhibit superior capacity to generalize in challenging nephropathological settings. These results further emphasize the need for models trained on large, diverse, multi-institutional datasets. The value of such a screening tool would have to be examined on an independent, ideally prospective dataset. We would expect our screening tool to alert the pathologist to the possibility of TMA on quite a substantial number of instances in mimickers, because such biopsies do show similarities to TMA and the differential diagnosis would always require careful integration of other histopathological and clinical findings by an expert nephropathologist.

5.1. Limitations

We would like to point out the limitations of our instance-level TMA classifier. As mentioned in Section 3.2.2, existing consensus criteria for TMA have only been established for transplant biopsies (Afrouzian et al., 2023). Consequently, in the absence of a consensus definition, the nephropathologists participating in this study (SS, AB, DD, SAMEGT, JUB) established a working definition and agreed upon a set of binary (present/absent) histopathological descriptors for the artery, arteriole, and glomerulus compartments. However, it should be pointed out that ground truth biopsy-level diagnoses were made under careful consideration of level sections, integration of all histopathological findings and the wider clinical context. It is also important to note that we are not aware of any negative histopathological finding precluding a diagnosis of TMA. Conceptually, these descriptors are rather vague, both by themselves and in combination. An example would be "platelet thrombus" as an accumulation of platelets. There is no agreement, about the minimal number (2, 3, more?) or proximity of platelets in a cluster defining such a "platelet thrombus". Consequently, creating a binary ML classifier TMA yes/no based on instance labels of individual tissue crops can only deliver a screening tool. Such a tool should make it easier for the nephropathologist to find the "needle in the haystack" for a diagnosis of TMA. Like all screening tools, it is the responsibility of the nephropathologist to carefully examine all positive instances within the context of other histopathological findings (including other modalities like electron microscopy) and the wider clinical context to either confirm or exclude a diagnosis of TMA.

6. Conclusion

We describe the development and performance of NePathTK, a uniquely robust and accurate multistaining segmentation and classification workflow trained and validated on renal biopsies with TMA and other vasculopathies. Considering the broad range of histopathological traits appearing in our data, our models can be used for the bulk segmentation of decisive tissue compartments Artery, Arteriole, Glomerulus, Cortex, Medulla, Capsule/Other on WSIs in clinical workflows and large renal biopsy registries encompassing different kidney pathologies. Our models unlock histopathological registries for compartment-wise analysis with a wide range of methodologies from the field of computer vision and machine learning. As an example, we provide a binary classifier as a screening tool for the histopathological diagnosis of TMA for expert nephropathologists.

Funding

The study was funded under the research projects:

NRRP project "BRIEF—Biorobotics Research and Innovation Engineering Facilities", Mission 4: "Istruzione e Ricerca", Component 2: "Dalla ricerca all'impresa", Investment 3.1: "Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione", CUP: J13C22000400007, funded by European Union—NextGenerationEU.

National Plan for Complementary Investments to the NRRP, project "D34H—Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care" (project code: PNC0000001), Spoke 2: "Multilayer platform to support the generation of the Patients' Digital Twin", CUP: B53C22006170001, funded by the Italian Ministry of University and Research.

European Union—NextGenerationEU – under the NRRP, Mission 4 Component 1 Investment 4.1 – Decree No. 118 (2nd March 2023) of Italian Ministry of University and Research – Concession Decree No. 2333 (22nd December 2023) of the Italian Ministry of University and Research, CUP: D93C23000450005, within the Italian National Programme PhD Program in Autonomous Systems (DAuSy).

Piano Operativo Salute, Traiettoria 3, project "Sistema Integrato Salute e Genetica Delle Patologie Cardio-Vascolari" (SISAGEN CARDIO), cod. T3-AN-18, CUP: D93C22000620004, funded by the Italian Ministry of Health.

CRediT authorship contribution statement

Nicola Altini: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Michela Prunella:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation. **Surya V. Seshan:** Validation, Data curation, Writing – review & editing. **Savino Sciascia:** Validation, Data curation, Writing – review & editing. **Antonella Barreca:** Validation, Data curation. **Alessandro Del Gobbo:** Validation, Data curation, Writing – review & editing. **Stefan Porubsky:** Validation, Data curation, Writing – review & editing. **Hien Van Nguyen:** Writing – review & editing. **Claudia Delprete:** Software, Methodology, Investigation, Writing – review & editing. **Berardino Prencipe:** Software, Methodology, Investigation, Writing – review & editing. **Deján Dobi:** Validation, Data curation, Writing – review & editing. **Daan P.C. van Doorn:** Validation, Data curation. **Sjoerd A.M.E.G. Timmermans:** Validation, Data curation, Writing – review & editing. **Pieter van Paassen:** Validation, Data curation, Writing – review & editing. **Vitoantonio Bevilacqua:** Writing – review & editing, Supervision, Investigation. **Jan Ulrich Becker:** Writing – review & editing, Writing – original draft, Validation, Supervision, Investigation, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare they have no competing interests.

Acknowledgements

Savino Sciascia and Jan Ulrich Becker are supported by the European Rare Kidney Disease Network (ERKNet). The work of Michela Prunella and Claudia Delprete is supported by the Italian National Program PhD Programme in Autonomous Systems (DAuSy).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.compmedimag.2025.102658.

Data availability

Data will be made available on request.

References

Afrouzian, M., Kozakowski, N., Liapis, H., Broecker, V., Truong, L., Avila-Casado, C., et al., 2023. Thrombotic microangiopathy in the renal allograft: results of the TMA banff working group consensus on pathologic diagnostic criteria. Transpl. Int J. Eur. Soc. Organ Transpl. 36, 11590.

Altini, N., Cascarano, G.D., Brunetti, A., Marino, F., Rocchetti, M.T., Matino, S., et al., 2020. Semantic segmentation framework for glomeruli detection and classification in kidney histological sections. Electronics 9 (3), 503.

Altini, N., Cascarano, G.D., Brunetti, A., De Feudis, I., Buongiorno, D., Rossini, M., et al., 2020. A deep learning instance segmentation approach for global glomerulosclerosis assessment in donor kidney biopsies. Electronics 9 (11), 1768.

Altini, N., Rossini, M., Turkevi-Nagy, S., Pesce, F., Pontrelli, P., Prencipe, B., et al., 2023. Performance and limitations of a supervised deep learning approach for the histopathological Oxford classification of glomeruli with IgA nephropathy. Comput. Methods Prog. Biomed. 242, 107814.

Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P. D., et al., 2017. QuPath: open source software for digital pathology image analysis. Sci. Rep. 7 (1), 16878.

Becker, J.U., Mayerich, D., Padmanabhan, M., Barratt, J., Ernst, A., Boor, P., et al., 2020. Artificial intelligence and machine learning in nephropathology. Kidney Int 98 (1), 65–75.

Berloco, F., Zaccaria, G.M., Altini, N., Colucci, S., Bevilacqua, V., 2025. A multimodal framework for assessing the link between pathomics, transcriptomics, and pancreatic cancer mutations. Comput. Med. Imaging Graph. 123, 102526.

Bouteldja, N., Klinkhammer, B.M., Bülow, R.D., Droste, P., Otten, S.W., Freifrau von Stillfried, S., et al., 2021. Deep Learning–Based segmentation and quantification in experimental kidney histopathology. J am soc nephrol. JASN 32 (1), 52–68.

Bueno, G., Fernandez-Carrobles, M.M., Gonzalez-Lopez, L., Deniz, O., 2020. Glomerulosclerosis identification in whole slide images using semantic segmentation. Comput. Methods Prog. Biomed. 184, 105273.

Cazzaniga, G., Rossi, M., Eccher, A., Girolami, I., L'Imperio, V., Van Nguyen, H., et al., 2023. Time for a full digital approach in nephropathology: a systematic review of current artificial intelligence applications and future directions. J. Nephrol. 37 (1), 65–76.

Chen L.C., Zhu Y., Papandreou G., Schroff F., Adam H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818).

Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R., 2022. Masked-attention Mask Transformer for Universal Image Segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1290–1299.

Geng Z., Guo M.H., Chen H., Li X., Wei K., Lin Z., 2021. Is Attention Better Than Matrix Decomposition?. *arXiv preprint arXiv:2109.04553*.

Ginley, B., Lutnick, B., Jen, K.Y., Fogo, A.B., Jain, S., Rosenberg, A., et al., 2019. Computational segmentation and classification of diabetic glomerulosclerosis. J. Am. Soc. Nephrol. 30 (10), 1953–1967.

Guo, M.H., Lu, C.Z., Hou, Q., Liu, Z., Cheng, M.M., Hu, S.M, 2022. SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation. Adv. Neural Inf. Process. Syst. 35, 1140–1156.

Haas, M., Seshan, S.V., Barisoni, L., Amann, K., Bajema, I.M., Becker, J.U., et al., 2020. Consensus definitions for glomerular lesions by light and electron microscopy: recommendations from a working group of the renal pathology society. Kidney Int 98 (5), 1120–1134.

He K., Zhang X., Ren S., Sun J., 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Hölscher, D.L., Bouteldja, N., Joodaki, M., Russo, M.L., Lan, Y.C., Sadr, A.V., et al., 2023. Next-Generation morphometry for pathomics-data mining in histopathology. Nat. Commun. 14 (1), 470.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Honolulu, HI, pp. 2261–2269.

Jaugey, A., Maréchal, E., Tarris, G., Paindavoine, M., Martin, L., Chabannes, M., et al., 2023. Deep learning automation of MEST-C classification in IgA nephropathy. Nephrol. Dial. Transpl. 38 (7), 1741–1751.

Jayapandian, C.P., Chen, Y., Janowczyk, A.R., Palmer, M.B., Cassol, C.A., Sekulic, M., et al., 2021. Development and evaluation of deep learning–based segmentation of histologic structures in the kidney cortex with multiple histologic stains. Kidney Int 99 (1), 86–101.

Jiang, L., Chen, W., Dong, B., Mei, K., Zhu, C., Liu, J., et al., 2021. A deep Learning-Based approach for glomeruli instance segmentation from multistained renal biopsy pathologic images. Am. J. Pathol. 191 (8), 1431–1441.

Kers, J., Bülow, R.D., Klinkhammer, B.M., Breimer, G.E., Fontana, F., Abiola, A.A., et al., 2022. Deep learning-based classification of kidney transplant pathology: a retrospective, multicentre, proof-of-concept study. Lancet Digit Health 4 (1), e18–e26.

Lin T.Y., Dollár P., Girshick R., He K., Hariharan B., Belongie S., 2017. Feature Pyramid Networks for Object Detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., Wei, F., Guo, B., 2022. Swin Transformer V2: Scaling Up Capacity and Resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 12009–12019.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al., 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 10012–10022.

Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A ConvNet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11976–11986.

Lutnick, B., Manthey, D., Becker, J.U., Ginley, B., Moos, K., Zuckerman, J.E., et al., 2022. A user-friendly tool for cloud-based whole slide image segmentation with examples from renal histopathology. Commun. Med 2 (1), 105.

Pfister, F., Amann, K., Daniel, C., Klewer, M., Büttner, A., Büttner-Herold, M., 2018. Characteristic morphological changes in anti-VEGF therapy-induced glomerular microangiopathy. Histopathology 73 (6), 990–1001.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer international publishing, Cham, pp. 234–241.

Roufosse, C., Simmonds, N., 2018. Clahsen-van Groningen M, haas M, henriksen KJ, horsfield C, et al. A 2018 reference guide to the banff classification of renal allograft pathology. Transplantation 102 (11), 1795.

Salvi, M., Mogetta, A., Gambella, A., Molinaro, L., Barreca, A., Papotti, M., et al., 2021. Automated assessment of glomerulosclerosis and tubular atrophy using deep learning. Comput. Med Imaging Graph 90, 101930.

Silva, J., Souza, L., Chagas, P., Calumby, R., Souza, B., Pontes, I., et al., 2022. Boundary-aware glomerulus segmentation: toward one-to-many stain generalization. Comput. Med Imaging Graph 100, 102104.

Tan, M., Le, Q., 2019. EfficientNet: rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning [Internet]. PMLR, pp. 6105–6114. In: ⟨https://proceedings.mlr.press/v97/tan19a.html⟩ (Available from).

Tan, M., Le, Q., 2021. EfficientNetV2: smaller models and faster training. Proceedings of the 38th International Conference on Machine Learning [Internet]. PMLR, pp. 10096–10106. In: ⟨https://proceedings.mlr.press/v139/tan21a.html⟩.

Vasiljević, J., Feuerhake, F., Wemmert, C., Lampert, T., 2021. Towards Histopathological Stain Invariance by Unsupervised Domain Augmentation using Generative Adversarial Networks. Neurocomputing 460, 277–291.

Vasiljević, J., Feuerhake, F., Wemmert, C., Lampert, T., 2023. HistoStarGAN: A unified approach to stain normalisation, stain transfer and stain invariant segmentation in renal histopathology. Knowl.-Based Syst. 277, 110780.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P., 2021. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. Adv. Neural Inf. Process. Syst. 34, 12077–12090.

Zare, S., Vo, H.Q., Altini, N., Bevilacqua, V., Rossini, M., Pesce, F., et al., 2025. Permutation-Invariant cascaded attentional set operator for computational nephropathology. Kidney360 6 (3), 441–450.