Glomerular basement membrane thickness estimation and stratification via active semi-supervised learning model

**American Journal of Nephrology**

| | |
|---|---|
| **Manuscript:** | AJN-2024-7-46/R1 RESUBMISSION |
| **Title:** | Glomerular basement membrane thickness estimation and stratification via active semi-supervised learning model |
| **Authors(s):** | Nico Curti (Co-author), Gianluca Carlini (Co-author), Sabrina Valente (Corresponding author), Enrico Giampieri (Corresponding Author), Alessandra Merlotti (Co-author), Daniel Remondini (Co-author), Gaetano La Manna (Co-author), Gastone Castellani (Co-author), Gianandrea Pasquinelli (Co-author) |
| **Keywords:** | artificial intelligence, Glomerular disorders, Nephropathy |
| **Type:** | Research Article |

*Research Article*

# *Glomerular basement membrane thickness estimation and stratification via active semi-supervised learning model*

Nico Curti [a,†], Gianluca Carlini [b,†], Sabrina Valente [c, d,*], Enrico Giampieri [c, d*], Alessandra Merlotti [a], Daniel Remondini[a], Gaetano La Manna[c,e], Gastone Castellani[c, d‡], Gianandrea Pasquinelli[c,f, ‡]

[a] Department of Physics and Astronomy, University of Bologna, 40127 Bologna (Italy)

[b] IRCCS Institute of Neurological Sciences of Bologna, Data Science and Bioinformatics Laboratory, 40139 Bologna (Italy)

[c] Department of Medical and Surgical Sciences, University of Bologna, 40138 Bologna (Italy)

[d] IRCCS Azienda Ospedaliero-Universitaria di Bologna, 40138 Bologna (Italy)

[e] Nephrology Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, 40138 Bologna (Italy)

[f] Pathology Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, 40138 Bologna (Italy)

Short Title: GBM thickness quantification by deep learning

Corresponding Authors:

Sabrina Valente

E-mail address: sabrina.valente2@unibo.it

Enrico Giampieri

E-Mail address: enrico.giampieri@unibo.it

[†] *These authors contributed equally to this study*

[‡] *These authors contributed equally to this study and share seniorship*

Keywords: computer-aided diagnosis; deep learning; image segmentation; glomerular diseases; GBM thickness

## Abstract

**Introduction:** The measure of Glomerular Basement Membrane (GBM) thickness is used as diagnostic criteria for kidney glomerular diseases. The GBM thickness measurement, a time-consuming task, is performed by expert pathologists on transmission electron microscopy (TEM) images, therefore, it is affected by subjectivity and reproducibility issues.

**Methods**: Here we introduce a fully automated pipeline for the GBM segmentation and successive thickness estimation, starting from TEM images. This method is based on an active semi-supervised learning training procedure of a convolutional neural network model. Starting from the areas automatically identified by the model, we provide a robust measurement of membrane thickness using pixels distance matrix and computer vision techniques. Using these values, we trained a machine learning model to automatically determine the GBM thickness. To verify the accuracy of the method, we compared the predicted results with the full iconographic materials and diagnostic record reports from 42 renal biopsies having normal-thick (n. 21), thin- (n. 10), thick-GBM (n. 11).

**Results**: The obtained segmentations were used for the automated estimation of GBM thickness via computer vision algorithms and compared with manual measurements, obtaining a correlation of Pearson's R2 of 0.85. The GBM thickness was stratified into 3 classes, namely normal, thin, thick with a 0.63 Matthews correlation coefficient and a 0.76 accuracy.

**Conclusion**: The proposed pipeline obtained state-of-the-art performance in GBM segmentation, proving its robustness under image variations, such as magnification, contrast, and complex geometrical shapes. Automated measures could assist clinicians in standard clinical practice speeding up routine procedures with high diagnostic accuracy.

## Introduction

In addition to histological stains and immunofluorescence analysis, electron microscopy plays a crucial role in diagnosing glomerular diseases[1,2]. Indeed, transmission electron microscopy (TEM) is an indispensable and recommended technique for the diagnostic assessment of medical renal biopsies[3].

Beyond specific diagnostic applications reviewed elsewhere[4-9], electron microscopy, thanks to its high magnification and resolution, is an essential tool for measuring the thickness of Glomerular Basement Membranes (GBM), an important criterion for diagnosing certain glomerular diseases. Thinning and thickening of the GBM are observed in genetic glomerular diseases, such as thin basement membrane disease and Alport syndrome[10,11], respectively. Diabetic nephropathy is characterized by diffuse thickening of otherwise normal-appearing GBM[12,13]. Thickening of the GBM due to enlargement of the lamina rara interna is observed in cases of endothelial injury, as seen in thrombotic microangiopathies and acute antibody-mediated graft rejections.

However, measuring the GBM is a laborious and time-consuming task, and accurate manual measurements require competent skills as well as the use of special strategies to yield reproducible results. Consequently, this activity is confined to selected centers, as there are few highly trained personnel worldwide in reading renal biopsies.

Manual measurement of membrane thickness could introduce biases related to the high inter- and intra- operator variability, the intra-operator errors in evaluations and no comparison between operators. The subjectivity of the operator plays a crucial role in the evaluation of the membrane thickness, potentially affecting its reliability. Moreover, there are no standardized measurement procedures, and the operator can only sample the membrane thickness in a series of arbitrary points, trying to obtain a reasonable representation of the global distribution of the values. A continuous, objective, and reproducible measure can only be addressed by automated segmentation methods. The feasibility of automated approaches relies on the availability of a robust set of annotated data for their training and validation. Automated solutions to address the membrane thickness evaluation were already proposed in the literature[14], aiming to standardize and speedup the measurement procedure. A first attempt was made by Ong et al. [15] in 1993, consisting in a semi-automated region-growing algorithm. Further semi-automated methods were proposed[16-18], relying on image processing techniques and requiring human intervention. Fully automated methods were initially proposed by Wu et al. [19] and Cao et al. [20]. The first method[19] uses a series of image processing techniques, such as image thresholding, edge detections, and image filtering to obtain the automated segmentation of the membrane and the measurement of GBM thickness. However, it is applied only to straight membrane segments and well contrasted images. The second technique[20] uses a Random Forest algorithm to achieve the automated segmentation of the membrane, classifying image pixels into background and foreground (membrane) classes. Nonetheless, the authors reported that their method is unable to segment the membrane in poor contrast conditions. Moreover, they did not provide any automated solution for the thickness estimation.

The introduction of Artificial Intelligence (AI) models poses the basis for novel approaches for automated segmentation algorithms. Convolutional Neural Networks (CNNs) are used in a wide variety of biomedical applications, usually achieving state-of-the-art results[21,22]. However, their implementation and usage require the collection of large amounts of annotated data, which is often unfeasible in medical applications and particularly time consuming for the clinicians. Active semi-supervised learning (ASSL) training strategy is showing promising results to overcome these kinds of issues, allowing a reasonably fast annotation of large datasets with a minimal effort by clinicians[23].

In this work we propose a CNN model for the automated segmentation of GBM membranes, trained on TEM images annotated via ASSL training strategy. Starting from the obtained segmentations a standard image processing pipeline was applied for the fully automated quantification of GBM thicknesses. Using these automated measures, we trained a machine learning model to stratify GBM thickness in a retrospective series of kidney biopsies from patients with glomerular diseases. This series included three groups of patients: patients without changes in GBM thickness (n. 21, control group), patients with thin GBM (n. 10), and patients with thick GBM (n. 11). By comparing the predicted GBM thickness[5,24-26] with the final diagnoses of the patients, the level of agreement obtained was very high (76%), suggesting the possibility of applying the machine learning model as an aid in clinical practice.

## Methods

### Selection of TEM images and patients

The TEM images were retrospectively collected from the files of the Pathology Unit at IRCCS Azienda Ospedaliero-Universitaria S. Orsola, Bologna. The data acquisition protocol was approved by the Local Ethics Committee, University-Hospital S. Orsola-Malpighi, Bologna (protocol n° RecoverEMO; 185/2020/Sper/AOUBo) according to the Helsinki Declaration.

TEM images belonged to 42 patients which were biopsied from 2020 to 2022 under routine nephrological examination. A total of 89 TEM images archived with each diagnostic report were used for the study. Patients were selected based on the diagnostic record, as follows, glomerular diseases without changes in GBM thickness (n. 21), with thin GBM (n. 10; 7 thin membrane disease, 3 unspecific ultrastructural lesions), with thick GBM (n. 11; 10 diabetic nephropathy, 1 chronic transplant glomerulopathy). In the images, the analyzed glomerular basement membranes had to belong to at least two different and non-sclerotic glomeruli; therefore, the two images are indeed representative and thus diagnostic[27].

Table 1 lists the main histopathological diagnostic findings of the enrolled patients.

### TEM image acquisition

TEM images were acquired during the routine clinical practice using a transmission electron microscope (Philips CM10, Eindhoven, Netherland) equipped with an Olympus digital camera, at 80 kV. For each patient, ten-to-fifteen digital images were recorded at either 4’800× or 13’500× and up to three images were associated with each diagnostic record and subsequently selected for the study; generally, these TEM images were those more meaningful from a diagnostic point of view.

### Data annotation

Two expert pathologists performed the manual annotation of a randomly selected subset of images. The annotation was performed by one expert and reviewed by the second one to improve data reliability. The manual annotation consists of a binary mask of the original image in which only the GBM areas are highlighted. This set of images with associated masks was used as the initial dataset for our ASSL training procedure.

Manual segmentation is very time consuming and pixel-wise annotation is difficult. We intentionally chose to minimize the number of manual annotations, relying on the robustness of the ASSL training strategy, leaving to the automated model the task of a pixel-perfect segmentation. An initial rough manual segmentation does not affect the effectiveness of ASSL training, as already shown by our previous works[23,28].

Starting from this small core set of annotated data, we performed multiple ASSL training rounds, progressively enlarging the number of annotated images, following the scheme already introduced[23]. We started from a core set of 15/89 images, and we divided each image into a series of disjointed 256×256 patches. For this task we decided to focus only on a subset of 69/89 images, keeping the remaining 20 images as an independent validation set for the evaluation of the final model. We repeated the active evaluation of the automated segmentations until the number of training patches reached at least 80% of available samples. For each ASSL round we split the available annotated data into train - test sets, quantifying the generalization capability of the model at each round.

The two expert pathologists performed the manual estimation of the GBM thickness for the entire set of images. For each image, the experts manually evaluated the thickness in (at least) 15 randomly selected portions of the GBM to achieve a sufficient statistic.

We would like to stress that the data annotation step is a mandatory task for the development of any machine learning model, but it is required only for the training of the model. Once the model was trained or it was released, the possibility to use it during clinical practice is feasible without any further effort by the clinicians. The only constraints related to its applicability remains bound to the hardware availability: all the simulations were performed using a 64-bit workstation machine (64 GB RAM memory and 1 CPU i9-10920X Intel®, with 12 cores, and a Quadro RTX 4000 NVIDIA® GPU). However, the described architecture was mandatory for the training step of the proposed pipeline, while less powerful machines could easily support the entire final prediction task.

## Segmentation model

In this work we used an EfficientNetB3[29] U-Net[30] CNN architecture as segmentation model, with the encoder pre-trained on the ImageNet dataset. EfficientNetB3 is a CNN model that is part of the EfficientNet family, designed for image classification tasks. It balances accuracy and computational efficiency by using a compound scaling method, which uniformly scales the network's depth, width, and resolution. Inserting EfficientNetB3 into a U-Net architecture involves using EfficientNetB3 as the encoder (downsampling path) in place of the traditional convolutional layers (ref. *Supplementary Material 1*). EfficientNetB3, pretrained on large datasets, extracts rich hierarchical features from input images. These features are passed through skip connections to the U-Net decoder (upsampling path), which reconstructs the image. This integration enhances U-Net's ability to capture complex

features while maintaining efficiency, making it particularly effective for tasks like image segmentation. The GBM areas are quite simple biological structures, with small heterogeneity in terms of possible texture, and well-defined morphological shapes. A CNN model able to segment them should not require many parameters, which could only lead to overfitting issues. In the preliminary phases of our work, we tested different model architectures, obtaining compatible results in terms of segmentation performances. For sake of brevity, in the current work we presented the best performances obtained during our trials, aiming to balance model complexity and computational time required for the training. Despite the EfficientNetB3 backbone is certainly a quite heavy architecture in terms of parameters (> 10 millions), its often preferable to other smaller architectures for its fast execution time. Furthermore, we want to remark that the ASSL training strategy requires multiple rounds of training, so it is crucial to balance the computational effort required.

We implemented the EfficientNetB3 U-Net CNN model using the Tensorflow Python library[31]. The model was trained for 100 epochs (at each ASSL training round), with Adam optimizer (learning rate of $10^{-4}$), and a batch size of 8 images.

We divided each image into a series of disjointed 256×256 patches. Only patches which included at least 1% of pixels related to GBM areas were used during the training. In each train-test split we ensured that the patches used for testing belonged to images not contained in the training set. At the end of each training the full images were rebuilt for the clinicians' validation, by simply concatenating all the predicted patches. We superimposed on the reconstructed image the mask predicted by our model (patch by patch) and the patch subdivision grid. We provided to the pathologists the reconstructed images, asking to validate each patch individually: the grid overlay guarantees the contextualization of the small patches into the global image, providing all the required information to the clinicians for a robust validation. All the patches which included missing or wrong predictions were marked as invalid. Only patches with a perfect segmentation of GBM areas were considered validated and used for the next round of training. A schematic representation of the pipeline is shown in Figure 1.

For each epoch we monitored two commonly used metrics in image segmentation tasks, namely, the IoU[32] metric and the F1-score[33]. The model was trained to minimize a combination of Dice loss[34] and Binary Focal loss[35] functions. The training was performed including an intensive data augmentation of the patches, including all possible rotations and horizontal and vertical flips.

## Thickness estimation

The estimation of GBM thickness requires a smooth segmentation of membrane shapes to prevent false positive measurements. Therefore, we split the available images into a set of overlapping 256×256 patches, with a stride of 64 pixels, in order to obtain a smooth prediction at the boundaries of the patches by averaging the prediction of the overlapping regions.

We further processed the obtained masks by applying a closing algorithm (with a circular kernel of shape 10×10) to fill small imperfections, and filtering on connected components with area lower than 1000 pixels. The estimation of kernel dimensions and area thresholding were performed in

agreement with the visual inspection by the expert clinicians, trying to maximize the segmentation quality.

We processed the validated masks using a custom image processing pipeline for the estimation of GBM thickness. We evaluated the distance between each pixel belonging to the areas identified by our model and the background. The pixels belonging to the central part of the mask acquired the maximum distance to the background, providing a measure of the half-thickness of the GBM. Furthermore, we applied an iterative erosion algorithm to extract the skeleton of the original mask. The obtained skeleton was used as a filter for the distance matrix, providing the medial axis of the segmented region, i.e., the distance values related to the set of pixels corresponding to the central line of the GBM and perpendicular to the medial axis. The proposed pipeline aims to mimic the orthogonal intercept - mean harmonic thickness method[36] preserving its quality of operator subjectivity reduction, while reducing the effort required by the clinicians for the measure acquisition. A schematic representation of the algorithm is shown in Figure 2.

The application of the proposed image processing pipeline guarantees the estimation of GBM thickness in each point, providing the global distribution of values along the entire structure. This measure is totally independent from human intervention and provides the sampling of the entire distribution of the thicknesses of the involved membrane. In this way, for each image, we can provide a statistically robust evaluation of GBM thickness, estimating the median of the distribution and the corresponding uncertainty given by the MAD (Median Absolute Deviation). The usage of rank-based statistics is more robust against possible small anomalous results of the segmentation.

For each image we collected the manual estimation of GBM thickness performed by expert pathologists. The manual estimation was performed on the raw images, i.e., without the segmentation mask overlayed, to avoid any possible bias in the measure. The first expert performed the measure on at least 15 random points on the membrane; the second expert validated the manual estimation guaranteeing the robustness of the measurements.

We compared the manual estimation with the automated values obtained by our fully automated pipeline using a linear regression. We filtered out the outliers of the predicted thickness distribution using a robust statistical filter, i.e., preserving all the thickness values within 1.48 MAD. We used a linear regression model for the estimation of the best fit between predicted and manual measurements, evaluating the correlation via the Pearson's $R^2$ correlation coefficient.

### Thickness classification and correlation with the diagnostic report

Using the thickness measures automatically predicted by our model, we developed a machine learning model for the correct stratification of the patients GBM. Using the GBM thickness values reported in the literature as a guide[5,24-26], the predicted values were grouped into three classes, hypothetically attributable to the conditions of thin-GBM, normal-GBM, and thick-GBM. The ground truth labeling was performed by an expert pathologist comparing the predicted thickness classification with the entire set of images available for each patient and the final diagnostic report.

From each membrane we estimated the entire distribution of the thickness values. We characterized the distributions measuring minimum, maximum, median, average, and standard deviation metrics,

providing for each patient a list of features. In the case of multiple images related to the same patients, the array of features was obtained by the average value of each item.

We standardized the features according to their medians and quantiles values, aiming to minimize the effect of possible outliers. We trained a support vector machine classifier for the automated prediction of the thickness-classes and the consequent stratification of the patients. A stratified 10-fold cross-validation was used to evaluate the model performance; for each split the data pre-processing was performed on the training fold only to not introduce cross-contamination. As term of comparison for the robustness of the classification a 2-fold stratified cross-validation was applied 100 times with different train-test splits.

Of the 42 patients, 10 were included in the thin membranes group, 21 in the normal membrane group, and 11 in the thick membrane group. To correctly monitor the unbalance of the classes, we used the Matthews Correlation Coefficient score for the quantification of the performances of our model. For sake of completeness, we also estimated the accuracy of the classifier model and related confusion matrix.

## Results

### Training with Active learning

We performed 6 rounds of ASSL training (including the first one) to obtain the target of ≥80% segmented patches. Starting from the initial set of 15 images, we generated 174 patches for the training and 52 for the testing of the model. At each round, the two expert pathologists evaluated the generated masks according to the criteria established in the Methods section: only the patches which satisfied all the criteria were used for the next round of the training. The model was trained with the same set of hyper-parameters, resetting the initial weights, at each round. We reported in Table 2 the information of the dataset used at each round of ASSL training.

### GBM segmentation

We show in Figure 3a the results obtained by our training procedure at the end of the 6$^{th}$ round of training. The output of the CNN model was binarized according to the threshold of 0.5 of confidence.

The trend of score metrics confirmed the choice of the model and the preliminary consideration about model complexity requirements: the GBM segmentation task is sufficiently simple to not require complex CNN models or a higher number of model parameters, since at the 3$^{rd}$ round the model has already reached a plateau on training performances after only 15 epochs.

The considered dataset involves images acquired with different magnification factors. The results obtained by the model did not show significant variability with respect to magnification factors; showing that the model is robust to heterogeneous data.

The final model was tested also on the 20 images excluded by the ASSL procedure. For each image an expert pathologist evaluated the entire segmentation performed by the model, considering an image as correctly segmented only if the entire portion of the GBM was identified by the model. The final evaluation of the expert clinician led to forward 17 out of 20 images, i.e., a performance of 85% of the proposed model.

### GBM thickness estimation

We processed each obtained mask according to our computer vision pipeline, removing possible spurious areas and maximizing the smoothness of the segmentation shapes. We applied the proposed thickness evaluation algorithm, and we compared the median of the distribution of values with the average of the manually annotated measurements, for each image. From the linear regression we obtained $y = 0.93x + 0.03$ between the two sets of measurements, with a Pearson's $R^2$ correlation coefficient of 0.85. We show in Figure 3b the resulting correlation.

The entire pipeline, starting from the raw image to the thickness estimation, requires around 1 minute on the considered computing architecture (64 GB RAM memory, CPU i9-10920X Intel®, 12 cores, and a Quadro RTX 4000 NVIDIA® GPU).

### Thickness classification

We processed the entire set of 89 images using the proposed pipeline for the thickness estimation. For each image a set of features describing the distribution of the thickness values was estimated and used to feed the machine learning model. In Figure 4 we show the results obtained by our classification model, expressed in terms of confusion matrix, accuracy score, and Matthew Correlation Coefficient (MCC).

Using a 10-fold cross validation, we obtained a performance of 76% accuracy and 0.63 MCC score for the stratification of the patients into the 3 available classes. We would like to stress that the reported performances were estimated using the thickness values extracted by our fully automated pipeline and quantified over a test set completely independent of the train set used for the tuning of the parameters.

### Correlation with the diagnostic report

The overall agreement between the predicted value and the final diagnosis was high, being equal to 76% using a 10-fold cross validation. Analogous results were obtained using the repeated 2-fold stratified cross-validation, achieving an average balanced accuracy of 73±7% and a corresponding MCC of 0.58±0.11. Analyzing the specific classes, we observed a concordance between the thin GBM class and the thin membrane diseases equal to 90%, there being a concordant prediction in 9/10 cases. The concordance between the thick GBM class and diabetic nephropathy was 64%, there being a concordant prediction in 7/11 cases. Analyzing the individual cases, in one case the biopsy was recovered from the paraffin block, a procedure which, as already described in the literature[37], determines a reduction in the thickness of the GBM; in the three remaining cases the GBMs had a normal thickness; this condition, which requires further investigation, could be attributable to various variables, such as diagnosis of early diabetes, optimal management and adherence to therapy, focal nephropathy. The concordance between the normal GBM class and the control cases was 76%, there being a concordant prediction in 16/21 cases. In this class the discordance was associated in one case with the recovery of the biopsy from the paraffin block, in two cases with inadequate fixation of the sample, in one case with glomerular ischemia, and in the remaining case with an immunologically related glomerular disease (IgA nephropathy).

## Discussion

We proposed a fully automated pipeline for the segmentation of the GBM and the estimation of its thickness. The proposed ASSL training strategy allowed us to minimize the required amount of manually annotated images and, consequently, the time required for the collection of the dataset. Starting from a core set of only 15 manually annotated images, we were able to achieve a complete training procedure with minimal effort by the pathologists, maximizing the efficiency of the training and the model performances. The use of transfer learning improved the time required for the model convergence and sped up the entire training procedure. To the authors' knowledge, the proposed pipeline represents the first approaches to address this task via deep learning models and a fully automated solution for the membrane thickness estimation.

The proposed computer vision pipeline was able to reproduce the manual thickness estimations with a correlation of 0.85 (Pearson's $R^2$ correlation coefficient), confirming its robustness in the evaluation. The proposed pipeline allows the monitoring of the entire GBM shape, avoiding geometrical issues, errors of measurements, and biases, whereas a human observer could only focus on a restricted set of feasible points. In this work we trained and tested our model on images with different magnifications and contrasts, aiming to maximize the robustness of our pipeline.

Furthermore, we proved the robustness of the proposed fully automated solution also in terms of patient stratification, reporting a performance of 76% accuracy. The possibility to estimate both the thickness score and the patient classification increases the potential usage of the proposed solution in clinical practice, drastically reducing the effort required by the clinicians. The importance of automated patient stratification, indeed, is linked to the possibility of classifying the patient and therefore directing clinical efforts towards the most urgent situations.

The main limit of our pipeline is related to the presence of complex geometrical shapes and low contrast areas in GBM images. We noticed that the main difficulties experienced by our segmentation model were found in proximity to dense GBM areas, in which it is difficult to define a clear GBM boundary. The same kind of issues were found in images with very low contrast in gray level distribution. The same issues are commonly experienced also by humans and consequently transmitted during the training to the automated algorithms. We carefully analyzed the sources of these issues (not reported in this work for sake of brevity) and they will be discussed in a future work.

A second limit of our work is related to the number of analyzed images. The use of image patches for the training of the model is a common data augmentation strategy, but it cannot overcome the limits of the available data. The efficiency and robustness of deep learning models relies on the availability of large and heterogeneous datasets. The data considered in our study involves single-center images acquired with the same TEM machine and identical digital camera; training with a multi-center dataset composed of images acquired with different microscopes and configuration parameters is mandatory to improve the robustness of the proposed pipeline before its deployment in clinical practice.

A third limitation of the study is represented by the pre-analytical phase. In fact, in 10/42 we obtained a discordant result which in four cases was attributable to the procedure for recovering the biopsy from the paraffin block [n. 2] or inadequate primary fixation [n. 2] as the biopsy had not been immersed in the fixative solution during the phase of sending the preparation for acceptance in the pathological anatomy. The greatest discordance was observed in samples from patients with diabetic nephropathy, where in three cases it was not possible to identify a plausible cause of the discordance. In these cases, which require further investigation, it is possible that insufficient sampling has led to the discrepancy.

In conclusion, despite the limit amount of data considered in the current study and the intrinsic issues due to the geometrical complexity of the shapes considered, the results obtained by the proposed fully automated pipeline could represent a valid starting point for future clinical

applications. We are currently working on the extension of the presented model, trying to increase the patient statistics both in terms of number of TEM scans and heterogeneity. The possibility of implementing and applying the proposed technology in the clinical practice, indeed, requires a greater number of cases, made by multiple prospective acquired on the same patient for a more robust estimation of the measurements, different microscopies to check the generalization capability of the segmentation model, and multiple pathological cases. Nevertheless, the results obtained by the proposed fully automated solution could represent a good starting point for future clinical applications.

## Acknowledgement

The authors would like to acknowledge Dr. Claudia Pisanti, who helped during the initial stage of this work, providing preliminary analysis of the segmentation models. The authors would like to acknowledge Elna Figuera Barbé, who contributed to the manual segmentation of the first batch of images for the kick-starting of active learning procedure and Dr. Elisa Albertini, who performed the revision of the clinical data.

## Statement of Ethics

The study was conducted according to the guidelines of the Declaration of Helsinki and was approved by Local Ethics Committee, University-Hospital S. Orsola-Malpighi, Bologna (protocol n° RecoverEMO; 185/2020/Sper/AOUBo).

## Consent Statement

This study was an retrospective single-center study and was approved by our local institutional review board. Informed consent was waived by the institutional review board due to the retrospective nature of the study.

## Conflict of Interest Statement

The authors have no conflicts of interest to declare.

## Funding Sources

This study was not supported by any sponsor or funder.

## Author Contributions

Nico Curti, Gianluca Carlini, Sabrina Valente, Enrico Giampieri and Alessandra Merlotti, performed study concept and design; Nico Curti, Gianluca Carlini, and Alessandra Merlotti performed development of methodology. All the authors contributing to the writing, review, and revision of the paper; Sabrina Valente and Gianandrea Pasquinelli provided acquisition and interpretation of data; Nico Curti and Gianluca Carlini provided statistical analysis; Daniel Remondini, Gaetano La Manna, Gastone Castellani and Gianandrea Pasquinelli provided material support. All authors read and approved the final paper.

## Data Availability Statement

The data that support the findings of this study are not publicly available due to restriction on clinical data due to EU, italian and regional regulations but are available from the corresponding author (EG or SV) on reasonable request. The pre-trained model and parameters used for the image

segmentation are available in the online repository, available at https://github.com/Nico-Curti/glomerular-basement-membrane.

## References


1. Walker PD, Cavallo T, Bonsib SM; Ad Hoc Committee on Renal Biopsy Guidelines of the Renal Pathology Society. Practice guidelines for the renal biopsy. Mod Pathol. 2004;17(12):1555-1563. doi: 10.1038/modpathol.3800239
2. Pavlisko EN, Howell DN. The continued vital role of electron microscopy in the diagnosis of renal disease/dysfunction. Ultrastruct Pathol. 2013;37(1):1-8. doi: 10.3109/01913123.2012.670025
3. Yamashita M, Lin MY, Hou J, et al. The Continuing Need for Electron Microscopy in Examination of Medical Renal Biopsies: Examples in Practice. Glomerular Dis. 2021;1(3):145-159. doi: 10.1159/000516831
4. Haas M. A reevaluation of routine electron microscopy in the examination of native renal biopsies. J Am Soc Nephrol. 1997;8(1):70-76. doi: 10.1681/ASN.V8170
5. Ramage IJ, Howatson AG, McColl JH, et al. Glomerular basement membrane thickness in children: a stereologic assessment. Kidney Int. 2002;62(3):895-900. doi: 10.1046/j.1523-1755.2002.00527.x
6. Roufosse C, Simmonds N, Clahsen-van Groningen M, et a. A 2018 Reference Guide to the Banff Classification of Renal Allograft Pathology. Transplantation 2018;102(11):1795-1814. doi: 10.1097/TP.0000000000002366
7. Haas M, Mirocha J. Early ultrastructural changes in renal allografts: correlation with antibody-mediated rejection and transplant glomerulopathy. Am J Transplant. 2011;11(10):2123-2131. doi: 10.1111/j.1600-6143.2011.03647.x
8. Weidemann F, Sanchez-Niño MD, Politei J, et al. Fibrosis: a key feature of Fabry disease with potential therapeutic implications. Orphanet J Rare Dis. 2013;8:116. doi: 10.1186/1750-1172-8-116
9. Santostefano M, Cappuccilli M, Gibertoni D, et al. Fabry Disease Nephropathy: Histological Changes With Nonclassical Mutations and Genetic Variants of Unknown Significance. Am J Kidney Dis. 2023;82(5):581-596.e0. doi: 10.1053/j.ajkd.2023.03.015
10. Foster K, Markowitz GS, D'Agati VD. Pathology of thin basement membrane nephropathy. Semin Nephrol. 2005;25(3):149-158. doi: 10.1016/j.semnephrol.2005.01.006
11. Haas M. Alport syndrome and thin glomerular basement membrane nephropathy: a practical approach to diagnosis. Arch Pathol Lab Med. 2009;133(2):224-232. doi: 10.5858/133.2.224

12. Marshall CB. Rethinking glomerular basement membrane thickening in diabetic nephropathy: adaptive or pathogenic? Am J Physiol Renal Physiol. 2016 1;311(5):F831-F843. doi: 10.1152/ajprenal.00313.2016

13. Mauer SM, Steffes MW, Ellis EN, et al. Structural-functional relationships in diabetic nephropathy. J Clin Invest. 1984;74(4):1143-1155. doi: 10.1172/JCI111523

14. Guo Y, Ashour AS, Sun B. A novel glomerular basement membrane segmentation using neutrsophic set and shearlet transform on microscopic images. Health Inf Sci Syst. 2017;5(1):15. doi: 10.1007/s13755-017-0036-7

15. Ong SH, Giam ST, Jayasooriah et al. Adaptive window-based tracking for the detection of membrane structures in kidney electron micrographs. Machine Vis. Apps. 1993;6:215–223. doi.10.1007/BF01212300

16. Kamenetsky I, Rangayyan RM, Benediktsson H. Analysis of the glomerular basement membrane in images of renal biopsies using the split-and-merge method: a pilot study. J Digit Imaging. 2010;23(4):463-474. doi: 10.1007/s10278-009-9233-5

17. Rangayyan RM, Kamenetsky I, Benediktsson H. Segmentation and analysis of the glomerular basement membrane in renal biopsy samples using active contours: a pilot study. J Digit Imaging. 2010;23(3):323-331. doi: 10.1007/s10278-009-9188-6

18. Wu HS, Dikman S, Gil J. A semi-automatic algorithm for measurement of basement membrane thickness in kidneys in electron microscopy images. Comput Methods Programs Biomed. 2010;97(3):223-231. doi: 10.1016/j.cmpb.2009.07.002

19. Wu HS, Dikman S. Segmentation and thickness measurement of glomerular basement membranes from electron microscopy images. J Electron Microsc (Tokyo). 2010;59(5):409-418. doi: 10.1093/jmicro/dfq060

20. Cao L, Lu Y, Li C, et al. Automatic Segmentation of Pathological Glomerular Basement Membrane in Transmission Electron Microscopy Images with Random Forest Stacks. Comput Math Methods Med. 2019;2019:1684218. doi: 10.1155/2019/1684218

21. Singh SP, Wang L, Gupta S, et al. 3D Deep Learning on Medical Images: A Review. Sensors (Basel). 2020;20(18):5097. doi: 10.3390/s20185097

22. Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. Evol Intell. 2022;15(1):1-22. doi: 10.1007/s12065-020-00540-3

23. Curti N, Merli Y, Zengarini C, et al. Effectiveness of Semi-Supervised Active Learning in Automated Wound Image Segmentation. Int J Mol Sci. 2022;24(1):706. doi: 10.3390/ijms24010706

24. Duan S, Sun L, Zhang C, et al. The thickness of glomerular basement membrane predicts complete remission in primary membranous nephropathy. Ren Fail. 2023;45(1):2179335. doi: 10.1080/0886022X.2023.2179335

25. Tyagi I, Agrawal U, Amitabh V, et al. Thickness of glomerular and tubular basement membranes in preclinical and clinical stages of diabetic nephropathy. Indian J Nephrol. 2008;18(2):64-69. doi: 10.4103/0971-4065.42336

26. Viana DL, Alladagbin DJ, Dos-Santos WLC, et al. A comparative study of human glomerular basement membrane thickness using direct measurement and orthogonal intercept methods. BMC Nephrol. 2022;23(1):23. doi: 10.1186/s12882-021-02634-1

27. Fogo AB, Kashgarian M. Diagnostic Atlas of Renal Pathology, 3rd Edition, Elsevier, 2021. ISBN: 9780323390538

28. Curti N, Giampieri E, Guaraldi F, et al. A Fully Automated Pipeline for a Robust Conjunctival Hyperemia Estimation. Applied Sciences. 2021;11(7):2978. doi: 10.3390/app11072978

29. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Published online 2019. doi: 10.48550/ARXIV.1905.11946

30. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Publidhed online 2015. doi: 10.48550/ARXIV.1505.04597

31. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Published online 2015. https://www.tensorflow.org/

32. Levandowsky M, Winter D. Distance between Sets. Nature. 1971;234(5323):34-35. doi: 10.1038/234034a0

33. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging. 2015;15:29. doi: 10.1186/s12880-015-0068-x

34. Sudre CH, Li W, Vercauteren T, et al. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support 2017; 2017:240-248. doi: 10.1007/978-3-319-67558-9_28

35. Lin TY, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection. Published online 2017. doi:10.48550/ARXIV.1708.02002

36. Dische FE. Measurement of glomerular basement membrane thickness and its application to the diagnosis of thin-membrane nephropathy. Arch Pathol Lab Med. 1992;116(1):43-49.

37. Valente S, Comai G, Malvi D, et al. Recovering histological sections for ultrastructural diagnosis of glomerular diseases through the pop-off technique. J Nephrol. 2021;34(6):2085-2092. doi: 10.1007/s40620-021-01043-9

## Figure Legends

**Fig. 1. Active semi-supervised learning pipeline.** We split each image into a series of independent patches, i.e., without overlapping. We fed our segmentation model using these patches and we re-built the entire image at the end of the prediction. We overlaid the computed mask and the split grid on the reconstructed image, asking the two expert pathologists to validate the obtained segmentations. If the obtained segmentation satisfied the established criteria, the obtained mask was used as a ground truth label in the next round of training; otherwise, it was discarded, and the corresponding patch kept as validation image in the next round of training.

**Fig. 2. Automated GBM thickness estimation**. Automated segmentation of GBM membrane via CNN model (Fig. 2 a); distance map estimation, i.e., the minimum distance of each pixel belonging to the membrane to the background area (Fig. 2 b); membrane skeleton extraction, i.e., the set of pixels belonging to the line passing through the center of the membrane (Fig. 2 c); the intersection of the skeleton pixels with the distance map provides the mathematically accurate estimation of membrane thickness, highlighted by magenta line in the figure.

**Fig. 3. Assessment of GBM thickness estimation**. Results obtained by the application of our automated pipeline. Comparison between a manual annotated segmentation and the automated result obtained using the model trained at the $6^{th}$ round of ASSL training (Fig. 3 a). Correlation plot between the thickness estimations performed manually by the two expert pathologists (y axis) and the values automated estimated by the proposed algorithm (x axis) (Fig. 3 b). For each measure, we report the average ± 1σ on the y axis and the median ± 1σ on the x axis.

**Fig. 4. Automated patient stratification.** Results obtained by the classification model for the stratification of the patients in terms of normal, thin, and thick membranes. Distribution of the patients according to the 3 classes identified by the expert clinician (Fig. 4 a). Distribution of the median thickness values automatically quantified by the proposed pipeline, split among the 3 classes of patients (Fig. 4 b). Confusion matrix and related performance scores (accuracy and Matthew Correlation Coefficient) obtained by the machine learning classifier trained on the thickness features extracted from the membrane segmentation (Fig. 4 c). The reported performances were estimated through a 10-fold cross validation procedure.

| Pz | Clinical notes | Histological / Ultrastructural diagnosis | GBM classification |
|---|---|---|---|
| 01 | Dialysis for end-stage renal failure. Hypertension and dysmetabolic syndrome. | **LM**: end-stage kidney with features suggestive of renovascular hypertension/ malignant hypertension.<br><br>**TEM**: glomerular ischemia. Absence of glomerular basement membrane alterations. | Normal |
| 02 | Worsening of renal function in transplant recipient. | **LM**: mild glomerular ischemia and minimal features of glomerulitis, in the context of mild chronic parenchymal modifications.<br><br>**TEM:** minor glomerular modifications.<br>Small mesangial deposits. | Normal |
| 03 | Worsening of renal function in transplant recipient (II transplantation). | **LM**: IFTA, grade II.<br><br>**TEM:** minor glomerular modifications. | Normal |
| 04 | Mild chronic renal failure. | **LM:** focal, segmental glomerular sclerosis, mesangial glomerulopathy and glomerular hypertrophy.<br><br>**TEM:** podocytopathy. | Normal |
| 05 | Acute renal failure in multiple myeloma (IgG/λ). | **LM:** acute tubular necrosis suspicious for Light Chain Cast Nephropathy.<br><br>**TEM:** minor glomerular lesions in acute tubular damage. | Normal |
| 06 | Mild chronic renal failure, systemic hypertension. | **LM:** mild glomerular hypertrophy, mild mesangial matrix | Normal |

| | | | |
|---|---|---|---|
| | | expansion and segmental glomerular sclerosis.<br><br>**TEM**: minor modifications. | |
| 07 | Focal segmental glomerulosclerosis. | **LM:** IgM nephropathy.<br><br>**TEM:** podocytopathy with mesangial deposits. | Normal |
| 08 | Acute renal failure associated with rhabdomyolysis. | **LM**: acute tubular damage and focal acute tubular necrosis.<br><br>**TEM**: minor glomerular modifications. Acute tubular necrosis. | Normal |
| 09 | Delayed graft function (II transplantation). | **LM:** mild aspecific parenchymal alterations.<br><br>**TEM:** minor glomerular lesions. | Normal |
| 10 | Worsening of renal function in transplant recipient, BKV+ | **LM**: BK virus-associated nephropathy (BKVAN).<br><br>**TEM:** minor glomerular modifications; interstitial BK nephritis. | Normal |
| 11 | Worsening of renal function, hypocomplementemia. | **LM:** monoclonal IgM mesangiopathic glomerulopathy.<br><br>**TEM:** podocytopathy with dense mesangial deposits. | Normal |
| 12 | Monoclonal gammopathy. | **LM:** suspicious for early tubular Ig deposition disease.<br><br>**TEM:** light chain deposition disease limited to tubular basement membrane. | Normal |

| | | | |
|---|---|---|---|
| 13 | Worsening of renal function in transplant recipient, DSA positive. | **LM:** acute antibody-mediated rejection.<br><br>**TEM:** glomerulitis and capillaritis. | Normal |
| 14 | Worsening of kidney function. | **LM:** diffuse acute tubulo-interstitial nephritis, with granulomas.<br><br>**TEM:** minimal glomerular alterations. Dense mesangial deposits. | Normal |
| 15 | Follow-up in AB0 incompatible kidney transplant. | **LM:** suspicious for active antibody-mediated rejection.<br><br>**TEM:** active antibody-mediated rejection. | Normal |
| 16 | Acute kidney injury on chronic kidney disease. | **LM:** end-stage kidney with cholesterinic arterial microembolism.<br><br>**TEM:** resolution phase of post-infectious glomerulonephritis in chronic vascular parenchymal damage. | Normal |
| 17 | Worsening of renal function in transplant recipient. | **LM:** focal proliferative mesangial glomerulonephritis and segmental glomerular sclerosis suggestive of recurrence of the previous disease (IgA nephropathy).<br><br>**TEM:** mesangiopathic glomerulonephritis with dense granular mesangial deposits. Recurrence of IgA nephropathy. | Normal |
| 18 | Chronic kidney disease, monoclonal gammopathy. History of prostatic and | **LM:** monoclonal immunoglobulin deposition disease with | Normal |

| | | | |
|---|---|---|---|
| | intestinal adenocarcinomas. | chronic severe tubulo-interstitial alterations.<br><br>**TEM:** dysmetabolic glomerulopathy and dominant tubulo-interstitial alterations related to monoclonal immunoglobulin deposition disease. | |
| 19 | Persistent microhematuria. | **LM:** IgA nephropathy, class II Haas.<br><br>**TEM:** mesangiopathic glomerulonephritis with dense granular mesangial deposits. IgA nephropathy. | Normal |
| 20 | Rapidly progressive renal failure in multiple myeloma IgG/λ. Suspicious for amyloidosis. | **LM**: renal AL/ λ amyloidosis.<br><br>**TEM**: no evidence of glomerular deposits. | Normal |
| 21 | Recurrent macrohematuria. | **LM**: IgA nephropathy, class III Haas.<br><br>**TEM:** mesangiopathic glomerulonephritis with dense mesangial and paramesangial deposits. IgA nephropathy. | Normal |
| 22 | Persistent microhematuria. Family history of renal disease. | **LM:** minor histologic modifications.<br><br>**TEM:** thin basement membrane disease. | Thin |
| 23 | Persistent microhematuria. | **LM:** mild mesangial glomerulopathy.<br><br>**TEM:** thin basement membrane disease. | Thin |
| 24 | Previous diagnosis of membranous glomerulonephritis of likely secondary origin. | **LM**: non significative glomerular alterations.<br><br>**TEM**: minor glomerular modifications. | Thin |

| | | | |
|---|---|---|---|
| 25 | Microhematuria.<br>Family history of nephropathy. | **LM:** mild focal mesangial glomerulopathy.<br><br>**TEM:** variable thickness of glomerular basement membrane, with thinning and layering. Minor glomerular lesions. | Thin |
| 26 | Chronic kidney disease. | **LM**: minor histological lesions.<br><br>**TEM**: thin basement membrane disease. | Thin |
| 27 | Mild chronic kidney disease.<br>Family history of nephropathy. | **LM:** minor histological lesions.<br><br>**TEM:** thin basement membrane disease. | Thin |
| 28 | Hypertension, and hearing loss. | **LM**: minor histological lesions with focal, segmental sclerosis.<br><br>**TEM:** thin basement membrane disease. | Thin |
| 29 | Suspected post-streptococcal glomerulonephritis. | **LM**: focal glomerulonephritis, with mesangial and membrane-proliferative patterns.<br><br>**TEM**: post-infectious glomerulonephritis outcomes in thin basement membrane disease**.** | Thin |
| 30 | Persistent microhematuria.<br>Family history of nephropathy. | **LM**: minor histological lesions.<br><br>**TEM**: thin basement membrane disease**.** | Thin |
| 31 | Proteinuria.<br>Family history of kidney disease. | **LM:** minor histological changes. | Thin |

| | | | |
|---|---|---|---|
| | | **TEM**: segmental glomerular basement membrane thinning. Structural glomerular basement membrane alterations not otherwise specified. | |
| 32 | Diabetes mellitus, hypertension and suspected Castleman syndrome. | **LM:** diabetic nephropathy, class IIb.<br><br>**TEM:** glomerular basement membrane thickening. Diffuse mesangial expansion. Dense mesangial and paramesangial deposits. | Thick |
| 33 | Diabetes mellitus. | **LM**: diabetic nephropathy, class III.<br><br>**TEM**: glomerular basement membrane thickening and mesangial expansion. | Thick |
| 34 | Acute kidney injury, diabetes mellitus and hypertension. Parathyroid adenoma. History of muco-epidermoid carcinoma. | **LM:** chronic focal pauci-immune glomerulonephritis, in diabetic nephropathy, class IIa and nephrocalcinosis.<br><br>**TEM:** mesangial expansion. Glomerular basement membrane thickening. | Thick |
| 35 | Kidney Transplant recipient. Diabetes mellitus, hypertension. | **LM:** diabetic nephropathy, class IIb.<br><br>**TEM:** diffuse mesangial expansion and glomerular basement membrane thickening. Diabetic nephropathy. | Thick |
| 36 | Diabetes mellitus. Rapidly progressive renal failure. | **LM**: diabetic nephropathy, class IIb<br><br>**TEM:** diffuse mesangial expansion and | Thick |

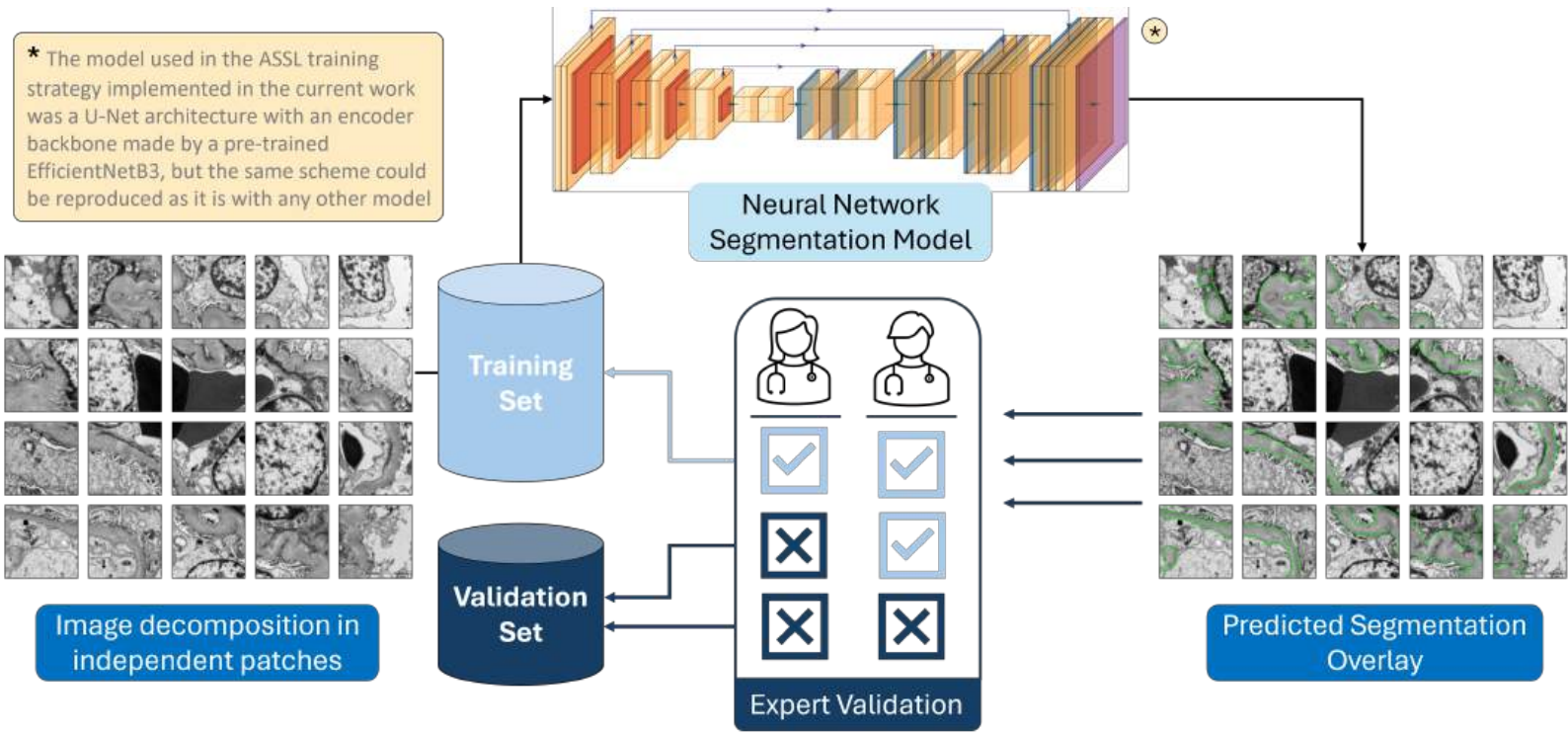
| | | | |
|---|---|---|---|
| | | glomerular basement membrane thickening. Diabetic nephropathy. | |
| 37 | Acute kidney injury on chronic kidney disease. Diabetes mellitus, FANS abuse. | **LM:** severe chronic parenchymal alterations.<br><br>**TEM**: diffuse mesangial expansion and glomerular basement membrane thickening. Diabetic nephropathy. | Thick |
| 38 | Diabetes mellitus, hypertension. | **LM:** diabetic nephropathy, class IIa<br><br>**TEM:** glomerular basement membrane thickening and arteriolar hyalinosis. Diabetic nephropathy. | Thick |
| 39 | Chronic kidney disease. Type 1 diabetes mellitus. | **LM**: diabetic nephropathy, class IV.<br><br>**TEM**: diabetic nodular glomerulosclerosis. | Thick |
| 40 | Diabetes mellitus, hypertension, drug-related hepatitis. | **LM:** chronic tubulo-interstitial nephritis and moderate arteriolosclerosis.<br><br>**TEM:** single, sclero-involved glomerulus. Chronic tubulo-interstitial damage. | Thick |
| 41 | Kidney Transplant recipient. | **LM**: chronic transplant glomerulopathy.<br><br>**TEM:** chronic parenchimal alterations. | Thick |
| 42 | Metabolic syndrome, IgG/k monoclonal gammopathy. History of metastatic renal carcinoma, under pharmacological treatment. | **LM:** complex glomerular alterations of difficult interpretation. Chronic moderate tubulo-interstitial damage. Severe arteriolar hyalinosis. | Thick |

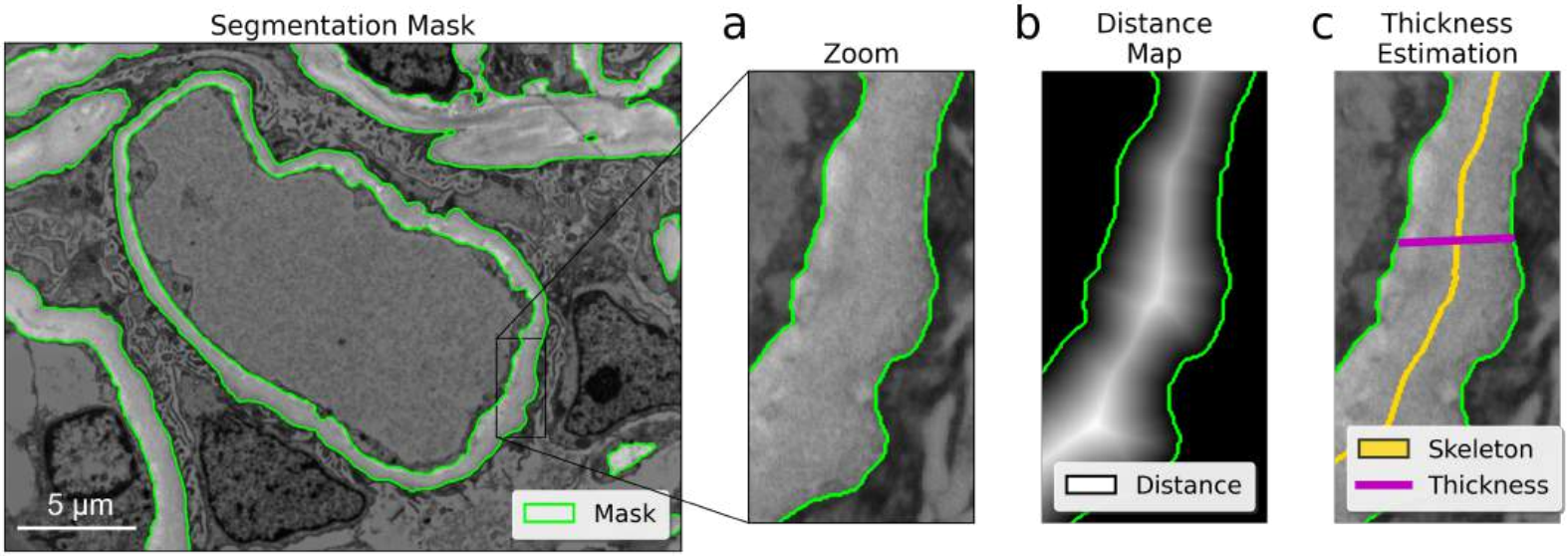
| | | | |
|---|---|---|---|
| | | **TEM:** microvascular damage and podocytopathy in a dysmetabolic context. | |

**Table 1.** Main histopathological diagnostic findings of the enrolled patients. Description of the dataset used in the proposed work. For each patient enrolled in the study, we list the main histopathological findings, including clinical and histological description of the sample, and final GBM classification, split into 3 main categories, namely normal, thin, and thick.

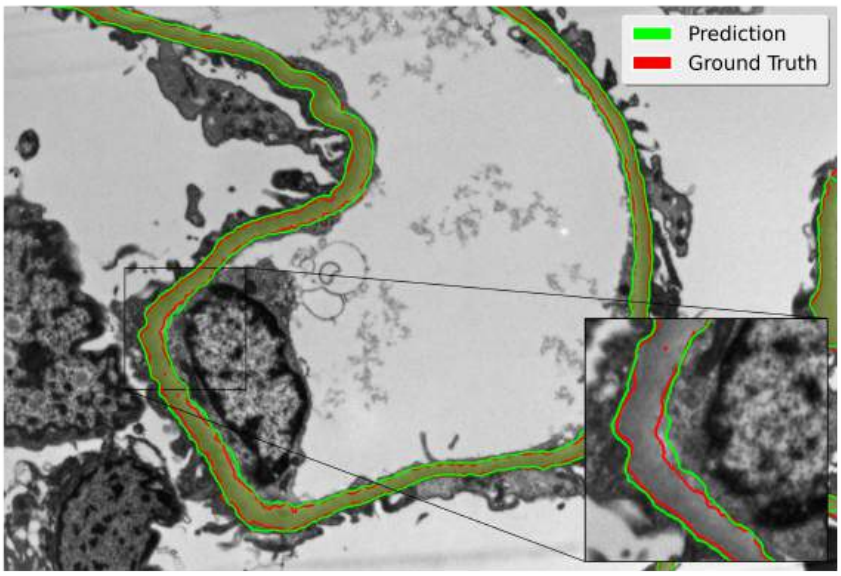
| | Round 0 | Round 1 | Round 2 | Round 3 | Round 4 | Round 5 |
|---|---|---|---|---|---|---|
| **N° training images** | 12 (17%) | 30 (70%) | 53 (77%) | 54 (78%) | 51 (74%) | 46 (67%) |
| **N° training patches** | 174 (16%) | 245 (23%) | 387 (37%) | 485 (46%) | 542 (52% | 583 (57%) |
| **N° testing images** | 3 (4%) | 8 (12%) | 10 (15%) | 12 (17%) | 15 (22%) | 21 (30%) |
| **N° testing patches** | 52 (5%) | 30 (3%) | 62 (6%) | 93 (9%) | 124 (12%) | 234 (22%) |
| **N° correct patches** | 275 (26%) | 449 (42%) | 578 (55%) | 666 (64%) | 817 (78%) | 845 (80%) |
| **IoU metrics** | 0.77 | 0.91 | 0.95 | 0.95 | 0.95 | 0.95 |
| **$F_1$ metric** | 0.87 | 0.95 | 0.98 | 0.98 | 0.97 | 0.97 |

**Table 2.** Results obtained by the EfficientNetB3 U-Net model using the ASSL training procedure at each round. We report the number of images used for the training, the number of patches extracted from the training images, the number of images used for the testing, the number of patches extracted from the testing images, the number of correctly segmented patches (according to the expert evaluation), and the metric scores achieved (on the test set) after 100 epochs for each round of training, respectively. The percentages of training and testing images are referred to the set of samples considered in the ASSL procedure, i.e., 69 images, while the percentages related to the patches are referred to the entire set of available patches, i.e., 1048 images. The number of correct patches was split between the train and test in the following round.

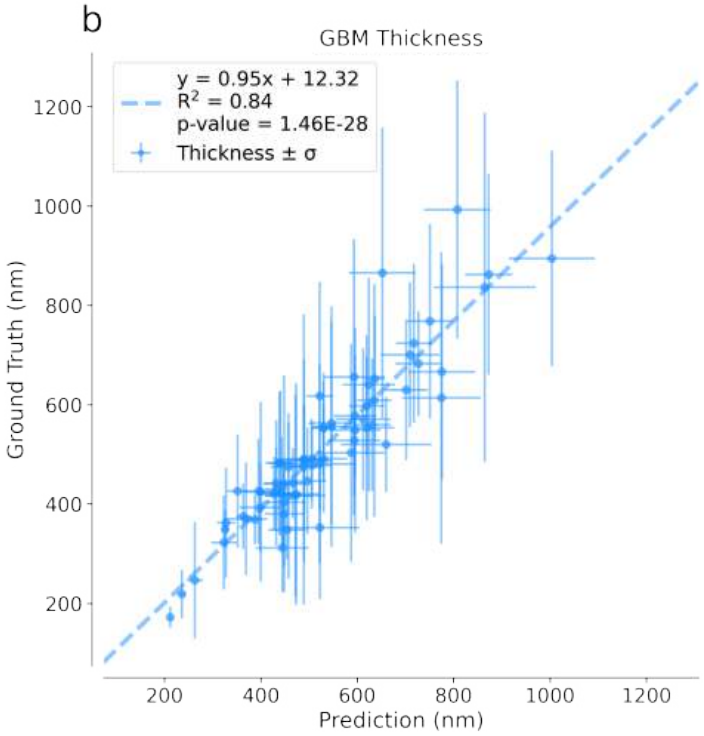

* The model used in the ASSL training strategy implemented in the current work was a U-Net architecture with an encoder backbone made by a pre-trained EfficientNetB3, but the same scheme could be reproduced as it is with any other model
Neural Network Segmentation Model
Training Set
Validation Set
Expert Validation
Image decomposition in independent patches
Predicted Segmentation Overlay

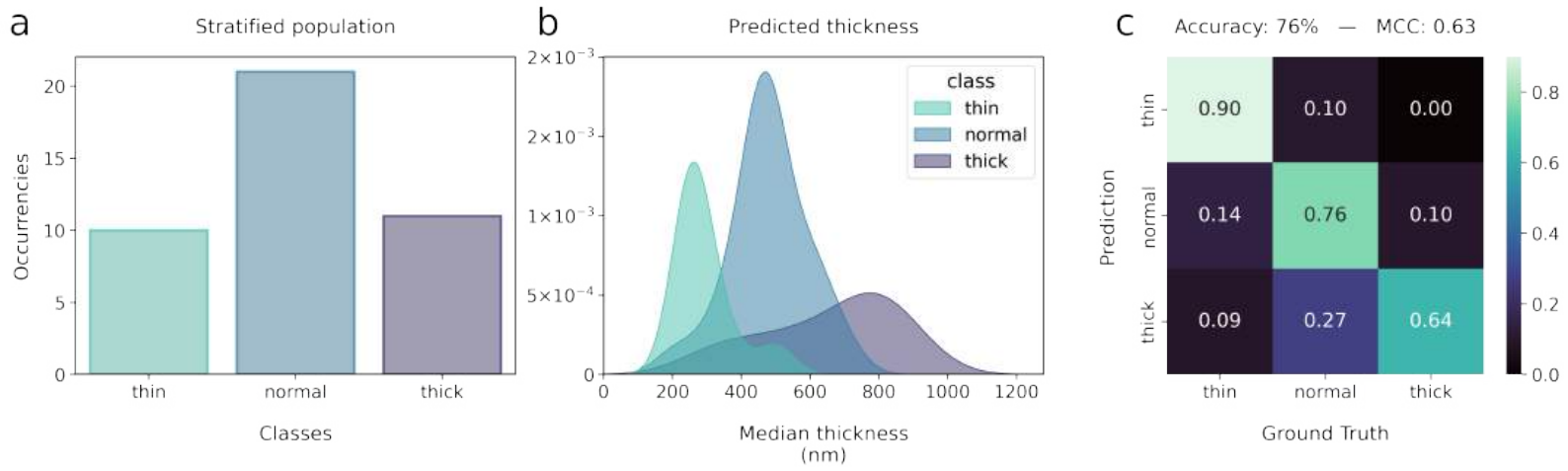

Segmentation Mask
5 µm
Mask
a
Zoom
b
Distance
Map
Distance
c
Thickness
Estimation
Skeleton
Thickness

a


Prediction
Ground Truth

b


GBM Thickness
y = 0.95x + 12.32
R² = 0.84
p-value = 1.46E-28
Thickness ± σ
Ground Truth (nm)
Prediction (nm)
200
400
600
800
1000
1200


a
Stratified population
Occurrencies
20
15
10
5
0
thin
normal
thick
Classes
b
Predicted thickness
$2\times10^{-3}$
$2\times10^{-3}$
$1\times10^{-3}$
$5\times10^{-4}$
0
class
thin
normal
thick
0
200
400
600
800
1000
1200
Median thickness
(nm)
c
Accuracy: 76% — MCC: 0.63
Prediction
thin
normal
thick
0.90
0.10
0.00
0.14
0.76
0.10
0.09
0.27
0.64
thin
normal
thick
Ground Truth
0.8
0.6
0.4
0.2
0.0

**Supplementary Material 1**

Schematics of the efficientnetb3_unet model used in the segmentation training. The design include all the parameters of each layers, their direct and skip connections.