Computerized Medical Imaging and Graphics 90 (2021) 101930

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

Automated assessment of glomerulosclerosis and tubular atrophy using deep learning

Massimo Salvi [a,*], Alessandro Mogetta [a], Alessandro Gambella [b], Luca Molinaro [c], Antonella Barreca [c], Mauro Papotti [d], Filippo Molinari [a]

[a] Politecnico di Torino, PoliToBIOMed Lab, Biolab, Department of Electronics and Telecommunications, Corso Duca degli Abruzzi 24, Turin, 10129, Italy
[b] Pathology Unit, Department of Medical Sciences, University of Turin, Via Santena 7, Turin, 10126, Italy
[c] A.O.U. Città della Salute e della Scienza Hospital, Division of Pathology, Corso Bramante 88, Turin, 10126, Italy
[d] University of Turin, Division of Pathology, Department of Oncology, Via Santena 7, Turin, 10126, Italy

ARTICLE INFO

Keywords:
Glomeruli segmentation
Tubular atrophy
Digital pathology
Kidney histology
Deep learning

ABSTRACT

In kidney transplantations, pathologists evaluate the architecture of both glomeruli, interstitium and tubules to assess the nephron status. An accurate assessment of glomerulosclerosis and tubular atrophy is crucial for determining kidney acceptance, which is currently based on the pathologists' histological evaluations on renal biopsies in addition to clinical data.

In this work, we present an automated algorithm, called RENTAG (Robust EvaluatioN of Tubular Atrophy & Glomerulosclerosis), for the segmentation and classification of glomerular and tubular structures in histopathological images. The proposed novel strategy combines the accuracy of a level-set with the semantic segmentation of convolutional neural networks to detect the glomeruli and tubules contours. In the TEST set, our method exhibited excellent performance in both glomeruli (dice score: 0.9529) and tubule (dice score: 0.9174) detection and outperformed all the compared methods.

To the best of our knowledge, the RENTAG algorithm is the first fully automated method capable of quantifying glomerulosclerosis and tubular atrophy in digital histological images. The developed software can be employed for the analysis of pre-transplantation biopsies to support the pathologists' diagnostic activity.

1. Introduction

Kidney transplantation proved to be the best treatment for patients with chronic end-stage renal diseases, showing a lower long-term mortality rate than dialysis. However, several factors – including the shortage of organ donors, and the increasing requests, especially of high-risk recipients - have increased the need for adequate stratifying criteria of recipients (Wolfe et al., 1999). In this critical setting, starting from the early 2000s, the United Network for Organ Sharing (UNOS) implemented the number of donor's kidneys with extended criteria donors (ECDs) (Port et al., 2002; Metzger et al., 2003). ECDs represent sub-optimal cadaveric donors with reduced functionality compared to standard criteria donors (SCDs) (Remuzzi and Perico, 1998), but they are still eligible for organ donation in a carefully selected cohort of patients. ECDs hold approximately 1/3 of the nephrons of an SCDs and therefore are more prone to present post-transplant functional issues (Rosengard et al., 2002). Then, it is crucial to establish whatever ECDs can support the long-term recipient's metabolic demands. To this purpose, a preimplantation biopsy is compelling to establish the nephron deterioration and the overall residual functionality and decide if the marginal kidneys are suitable for transplantation and eventually if a dual renal transplant is required (Mazzucco et al., 2010; Hassan and Halawa, 2015).

Biopsies are analyzed in the Pathology laboratory, where several procedures for biopsy preparation (e.g., frozen section, Serra fixation) and staining (e.g., trichrome, hematoxylin and eosin, PAS) are available, based on laboratory experience. One of the most widespread and useful staining for nephron evaluation is the PAS (i.e., Periodic Acid- Schiff reactive) because it highlights the polysaccharides and the glomerular basement membranes.

The pathologist analysis assessed the nephron status using a damage grading system of several functional structures. To standardize this evaluation, a specific score system, called "Karpinski score", has been proposed by an international group of pathologists (Karpinski et al.,

* Corresponding author.
E-mail address: massimo.salvi@polito.it (M. Salvi).

https://doi.org/10.1016/j.compmedimag.2021.101930
Received 20 July 2020; Received in revised form 20 February 2021; Accepted 21 April 2021
Available online 2 May 2021

1999). The score evaluates histologic injuries of crucial structure for kidney function in a two-step procedure. First, the pathologist assesses the percentages of sclerotic glomeruli, the tubular atrophy, the interstitial fibrosis, and the arterial and arteriolar narrowing. Then, each variable's evaluation is summarized the evaluation in a four-grade score, ranging from grade 0 (the absence of injury) to grade 3 (marked injuries). Thanks to this score, pathologists provide an easy and readable picture of kidney functionality, allowing the multidisciplinary transplant team to define the best management for these kidneys. Indeed, based on Karpinski scores, the equip decides if the kidneys are functionally adequate for a single transplant procedure, or if a double-transplant is required.

This paper is focused on the automated detection and classification of the glomerular and tubular structures and injuries within the renal tissue. Glomeruli are composed of networks of capillaries to filter waste substances from the bloodstream. Then, renal tubules reabsorbed water and the useful molecules accidentally filtered by glomeruli and convey and expel the wasted ones through urine. Glomerulosclerosis is referred to as nonspecific histological damage partially or entirely affecting the glomerulus, resulting in its very last form in the glomeruli capillary collapse (Jefferson and Shankland, 2014). To quantify glomerulosclerosis, pathologists assessed the percentage of sclerotic glomeruli among all the glomerular structures identified in each renal biopsy (Bueno et al., 2020). Regarding tubules, their pattern of injury is called tubular atrophy, and consist of the thickening of their walls and the reduction of their lumens. Tubular atrophy is a threatening condition that can lead to tissue necrosis and is evaluated by pathologists defining the percentage of atrophic tubules compared to all the tubules within the biopsy.

Because of the large number of structures to be manually detected, counted, and analyzed, kidney biopsies evaluations are generally time-consuming for pathologists, with a potential but relevant intra- and inter-observer variability (Bukowy et al., 2018).

Current research is focused on developing fully automated methods for quantitative histological analysis to create accurate tools with high reproducibility (Janowczyk and Madabhushi, 2016; Salvi et al., 2020a; Sharma et al., 2017). These approaches can provide a precise measure for analysis and reduce inter- and intra-observer variability in the measurement of cellular structures, increasing the reproducibility of the results (Litjens et al., 2016; Belhomme et al., 2015; Jha and Dutta, 2019). Automated glomerular detection is crucial in developing rapid and robust diagnostic tools to quantify glomerulosclerosis in whole-slide images (WSI) (Gadermayr et al., 2019). Segmentation and classification of glomerular structures is a demanding task and presents three main difficulties. Firstly, glomeruli show variable shapes, dimensions, and internal architecture, especially in pathological conditions, which makes their full detection challenging within a WSI. Secondly, WSI's large size requires an efficient and rapid method to detect all the glomerular structures inside the image. Finally, since multiple microscope slides may present differences in staining of the renal tissue, the algorithm of glomerular segmentation and classification should be robust against staining variations of the WSIs.

Similarly, tubules segmentation is a challenging task because of tubules variability in shape, dimension, and orientation. Normal and abnormal renal tubules are very close to each other, and the main difference between healthy and pathological structures is the increased thickness of the basal membrane and the shrinkage of the epithelial layer. Furthermore, a WSI may contain a considerable number of tubules, and the correct detection of all these structures is essential for a quantitative assessment of the degree of tubular atrophy.

In this paper, a novel method for the assessment of glomerulosclerosis and tubular atrophy is presented. This proposed algorithm combines deep learning with cellular structures detection to obtain an accurate segmentation and injuries classification of kidney glomeruli and tubules. Our approach is able to accurately detect the contours of kidney structures even in the presence of severe atrophy. Our algorithm is validated on 830 PAS stained images and outperforms all the state-of-art methods (Gallego et al., 2018; Kannan et al., 2019; Kawazoe et al., 2018; Bueno et al., 2020; Altini et al., 2020). Implementing this algorithm in pathology departments' daily clinical practice could represent a breakthrough in pre-transplant kidney assessment. The algorithm will standardize the Karpinski score evaluation, improving and accelerating its assessment to develop overall clinical management of patients awaiting a kidney transplant. Furthermore, the algorithm could also assist pathologists with little experience approaching this field. The rest of the manuscript is organized as follows: Section 2 presents an overview of the current approach for glomeruli segmentation and classification; Section 3 provides an exhaustive description of the proposed algorithm; Sections 4 and 5 report and discuss the experimental results.

2. Related works

Recently, several automated methods were proposed for quantitative assessment of kidney histological slides (Gallego et al., 2018; Kawazoe et al., 2018; Kannan et al., 2019; Bueno et al., 2020; Altini et al., 2020). Different approaches have been used to detect glomerular structures automatically, but there is still a lack of studies about the detection of tubular structures. Existing evidence suggests using Machine Learning for glomeruli detection, classification, or segmentation and specifically "deep learning" algorithms such as Convolutional Neural Networks (CNNs).

Gallego et al. (2018) employed a pre-trained AlexNet to detect normal and sclerotic glomeruli. The network was trained using cropped images containing Glomerulus and Non-glomerulus regions as inputs, and corresponding labels as outputs. Glomeruli detection and classification are then obtained, passing a sliding window operator over WSIs and applying the CNN prediction to the image block under analysis. Although this approach allows a correct detection of the glomerular regions, it is computational time consuming and provides a suboptimal classification result which can lead to an incorrect evaluation of the degree of glomerulosclerosis. Kannan et al. (2019) implemented an Inception architecture trained on patches extracted from trichrome stained images and labeled into three categories (no glomerulus, healthy glomerulus, and sclerotic glomerulus). By applying the CNN across an entire image, a heatmap was generated that indicated the probability of each pixel being a glomerulus (probability map). Then, a threshold was applied to the heatmap to detect healthy and sclerotic glomeruli. Being a patch-based classification approach, this method is time-consuming, and it also provides sub-optimal segmentation results.

Kawazoe et al. (2018) proposed a different approach based on the use of Faster R-CNN for glomeruli detection in WSI. This method took a cropped image obtained from a WSI using a sliding window and localized each glomerulus by drawing the corresponding bounding box. This approach was able to detect both healthy and sclerotic glomeruli but, being based on an object detector tool, it provided an over-segmentation of the glomeruli, which may cause incorrect detection of structures close to each other. Bueno et al. (2020) proposed a semantic segmentation of glomerular structures using an encoder-decoder network based on the VGG16 model. Pixel level segmentation is performed considering a two-class problem (non-glomerular vs. glomerular structures), and the segmentation results are then classified by an AlexNet to divide the glomeruli into normal and sclerotic. Using a two-step approach (segmentation followed by classification), all the errors of the segmentation network directly affect the classification one, which may cause the propagation of the misclassification error. Finally, Altini et al. (2020) proposed an approach based on DeepLab semantic segmentation network to obtain a pixel-wise classification of healthy and sclerotic glomeruli. The network was trained giving as input the image and the corresponding encoded mask labeled into 3 classes (healthy glomeruli, sclerotic glomeruli, background). The segmentation outputs were then post-processed using morphological operators followed by clustering to divide touching objects. This strategy provides a correct segmentation of

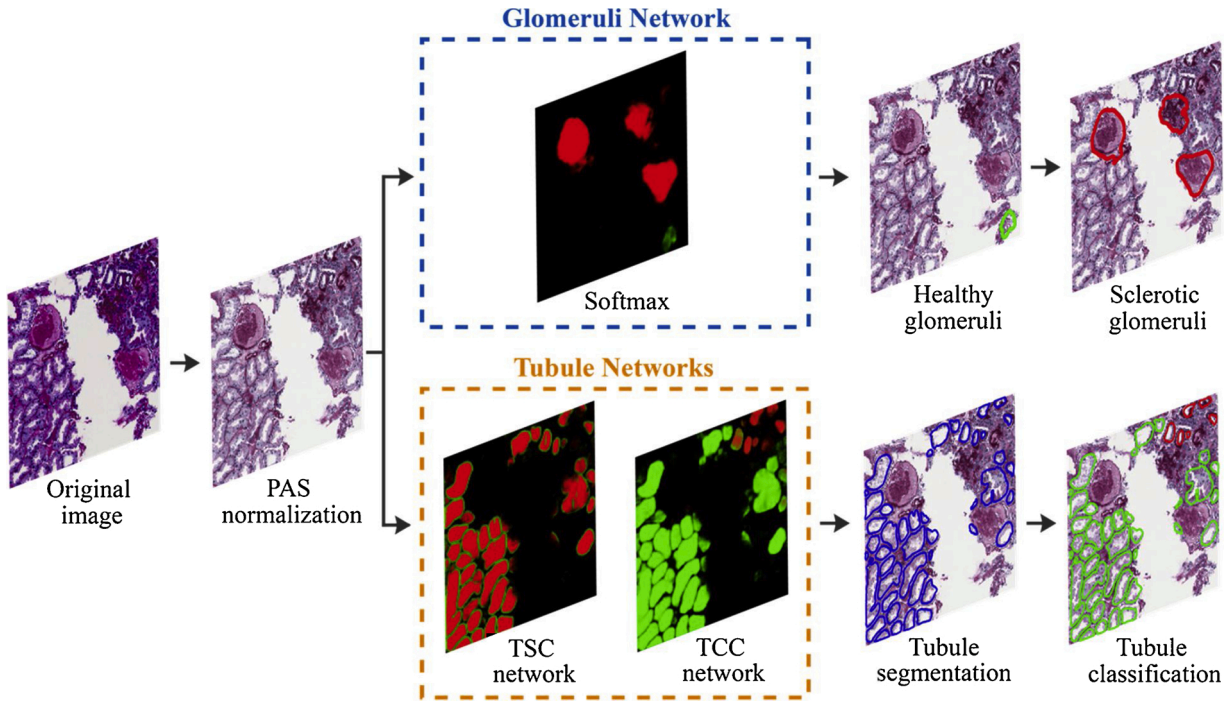

Fig. 1. Flowchart of the RENTAG algorithm for glomeruli & tubule segmentation and classification. Starting from the RGB image, a PAS normalization method is employed. Glomeruli are detected using a deep-learning network and ad-hoc post-processing. Kidney tubules are segmented and classified using two different deep networks (TSC and TCC).

healthy glomerular structures while the global performance deteriorates during the detection of sclerotic glomeruli (where the segmentation task is more challenging). To the best of our knowledge, no automated solution was proposed so far for quantifying tubular atrophy in renal histological images.

3. Materials and methods

In this paper, an automated method called RENTAG (Robust EvaluatioN of Tubular Atrophy & Glomerulosclerosis) is presented. The RENTAG algorithm is a deep learning-based method for the segmentation and classification of glomeruli and tubule in PAS stained images. The flowchart of the proposed method is shown in Fig. 1. The algorithm consists of three modules: PAS normalization, glomerulosclerosis assessment, and tubular atrophy quantification. In the following sections, a detailed description of the algorithm is provided.

3.1. Human kidney histology

The WSIs of kidney biopsy specimens of 83 patients (median age 67 years, range 38–89 years) were used for this work. In particular, 61 WSIs were used for glomerulosclerosis assessment while 22 WSIs were employed for tubular atrophy quantification. According to the Pathology Laboratory Protocol for kidney transplant biopsies, the tissues were collected by core needle biopsy, Serra-fixed, rapidly processed using the microwave-enhanced routinely tissue processor or the dedicated microwave tissue processor LOGOS J (Milestone, Bergamo, Italy), and paraffin-embedded, serially sectioned to 5 μm, mounted onto adhesive slides, and stained with PAS. All biopsy samples were collected at the Division of Pathology, AOU Città della Salute e della Scienza Hospital, Turin, Italy and were anonymized by a pathology staff member not involved in the study, before any further analysis was started. Digital images were obtained with a Hamamatsu NanoZoomer S210 Digital slide scanner providing a magnification of x100 (conversion factor: 0.934 μm/pixel). Ten images with a fixed dimension of 512 × 512 pixels were extracted from each patient (n = 83), for a total of 830 images.

Table 1
Dataset composition.

Dataset	Subset	#Patients	#Images	#Annotated shapes
Glomeruli	TRAIN	50	500	473
	TEST	11	110	114
Tubule	TRAIN	20	200	10,218
	TEST	2	20	868

After consensus, manual annotations of glomeruli and tubule contours were generated by two of us (LM and AG), for a total of 11,673 shapes. Table 1 shows the overall dataset composition.

3.2. PAS normalization

The first preprocessing step of the proposed pipeline is PAS normalization. Starting from the original RGB image of the specimen, the RENTAG algorithm performs the same stain normalization method that we developed in our previous work (Salvi et al., 2020b). Our color normalization strategy is based on stain separation. The contribution of the individual dye is isolated to alter the original image according to the color distribution of the reference image. As a result, all normalized images have their intensity distribution mapped to match the color distribution of the reference image (Fig. 2). PAS normalization is a crucial step as the appearance of histological stains often suffers from large variability due to the chemical reaction of the dye used during staining and the operator's ability. In this context, the standardization of the color appearance plays a crucial role in the development of deep learning solutions for quantitative analysis of histopathological images (Chen et al., 2020; Salvi et al., 2020a; Anghel et al., 2019).

3.3. Glomerulosclerosis assessment

3.3.1. CNN semantic segmentation

A deep CNN is employed to detect all the glomerular structures using Keras framework. In particular, a UNET network's architecture with

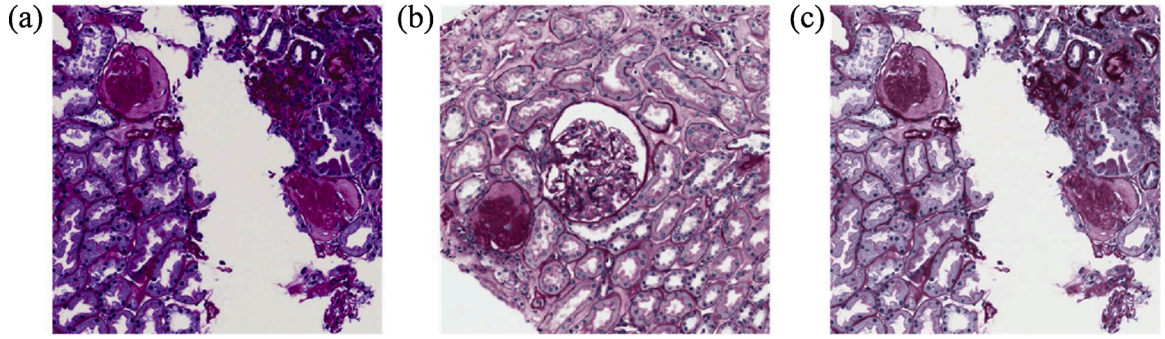

Fig. 2. Stain normalization process of the proposed algorithm. (a) original source image, (b) reference image, (c) normalized source image.

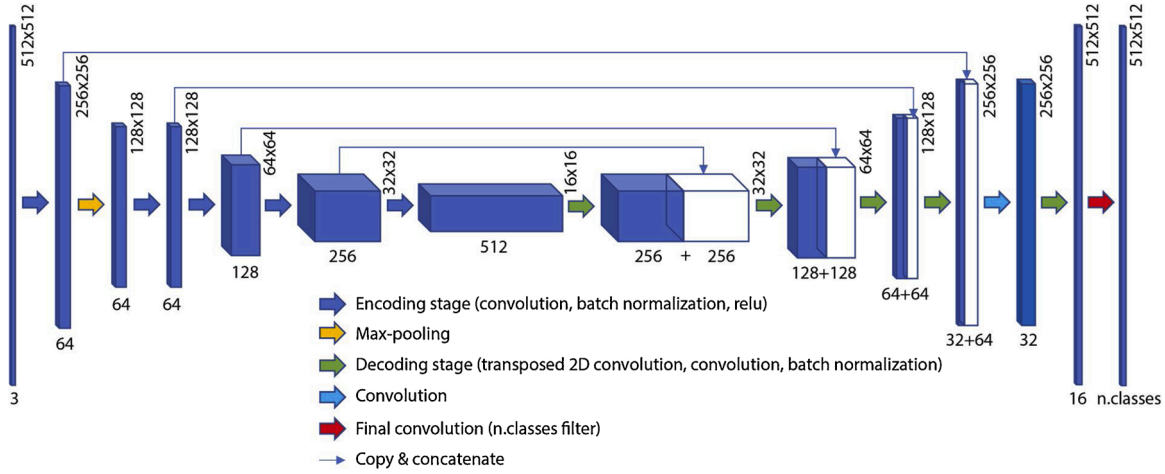

Fig. 3. Architecture of the deep network employed to perform glomeruli segmentation. A UNET with ResNet34 backbone was implemented using Keras framework.

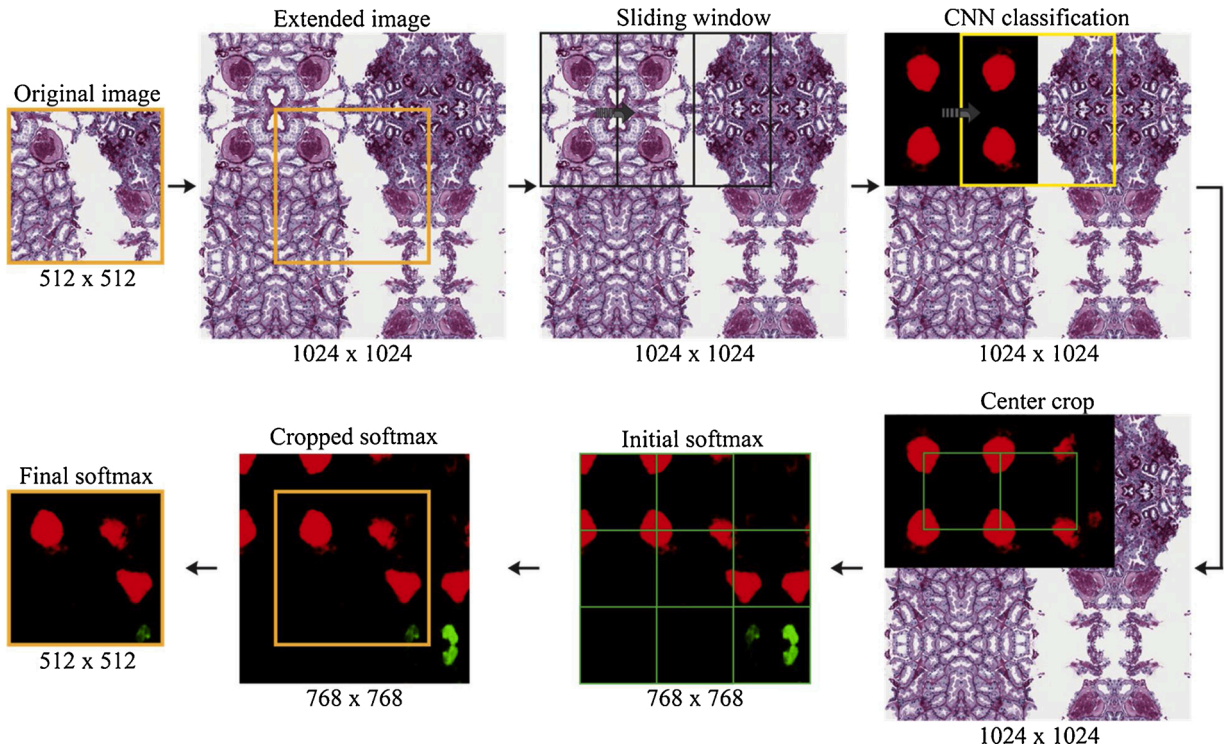

Fig. 4. Procedure for the creation of the final CNN softmax. The original image is mirrored around the boundaries to obtain the extended image. Then, a sliding window approach is employed to classify each patch, and only the center of each prediction is kept to build the final softmax.

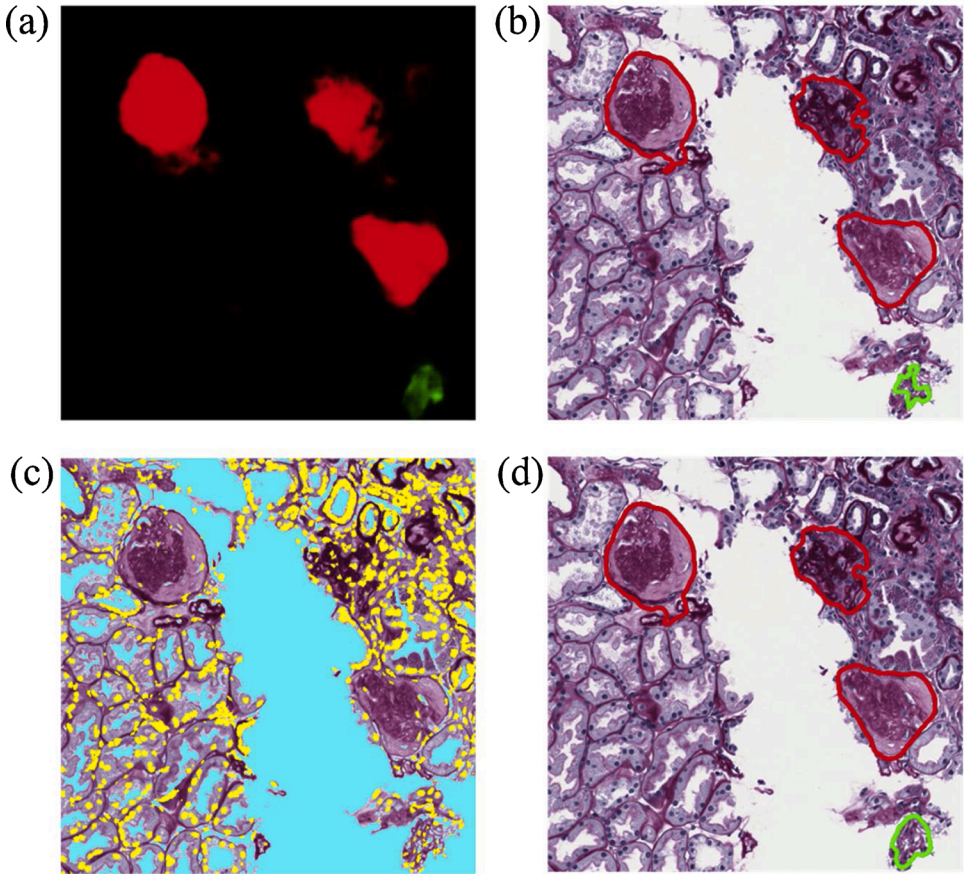

Fig. 5. Glomeruli detection and classification. (a) Softmax obtained using the deep networks (red: sclerotic glomeruli, green: healthy glomeruli), (b) Raw glomeruli detection, (c) Cellular structure detection (cyan: lumen, yellow: nuclei), (d) Final segmentation result of the proposed method. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ResNet34 backbone is adopted to perform semantic segmentation (Fig. 3). This network is composed of an encoder and a decoder structure. The encoder network, based on the ResNet34 architecture, downsamples the spatial resolution of the input image to obtain a low-resolution feature mapping. The decoder network is obtained by mirroring the encoding part using transposed convolutions. The aim is to semantically project the discriminating characteristics (lower resolution) learned by the encoder on the pixel space (higher resolution) to obtain a dense classification (He et al., 2016). This results in a full-resolution segmentation map where each pixel is labeled in the most likely class using a softmax layer.

Our deep convolutional network is based on residual blocks, which consists of a sequence of layers with skip connections between input and output of each block (He et al., 2016). The input layer is changed to fit the size of the input images ($512 \times 512 \times 3$). Each block of the encoding stage is composed of a Convolutional layer and a Batch Normalization layer. The decoding stage is obtained through a transposed convolution followed by a Convolution layer and a Batch Normalization layer. At the end of each stage, we used the Rectified Linear Activation Function (ReLU) between the input and the output.

The entire network is trained on a three-class problem, giving the 512×512 RGB images as input and the corresponding labeled mask as the target. To detect glomerular structures, pixels in the image are labeled into three different classes depending on whether they belong to healthy glomeruli, sclerotic glomeruli or other parts of the tissue. To solve the problem of class imbalance, our network's loss function is class-weighted by taking into account how frequently a class occurs in the training set. This means that the least represented classes (healthy and sclerotic glomeruli) will have a greater contribution than the more represented one (background) during the weight update. The class weight is calculated as follows:

$$f_{classX} = \frac{1}{N} \sum_{i=1}^{N} \frac{pixel_{classX}}{area(I)}, X = 1, 2, 3 \quad (1)$$

$$class_{weightX} = \frac{median(f_{class1}, f_{class2}, f_{class3})}{f_{classX}} \quad (2)$$

where I represents the current image, N is the total number of images, and f_{classX} is the class frequency of the generic class X.

The encoding part of our network is pre-trained on the 2012 ILSVRC imageNet Dataset (Krizhevsky et al., 2012). During the training process, only the decoder weights are updated, while the encoder weights are set to non-trainable. This strategy allows transferring the knowledge acquired from a previous task (ImageNet) to solve a new problem (glomeruli segmentation). This approach is useful both to overcome the problem of small datasets and to reduce the training time. Moreover, since histopathological images do not have a canonical orientation, on-the-fly data augmentation (flipping, rotating, scaling) is applied to increase the robustness of our network and avoid overfitting (Tellez et al., 2019; Yan et al., 2020). During deep network training, 10 % of the training images were selected as the validation set. The UNET is trained with a mini-batch size of 32 and an initial learning rate of 10^{-3}. Categorical cross-entropy and the Adam optimizer are employed as a loss and optimization function, respectively. Finally, the maximum number of epochs is set to 30, with a validation patience of 10 epochs for early stopping of the training process. The total training time was 14 h on a dedicated workstation equipped with a GeForce 1080Ti, 64 GB of RAM, and a 3.8 GHz ten-core CPU.

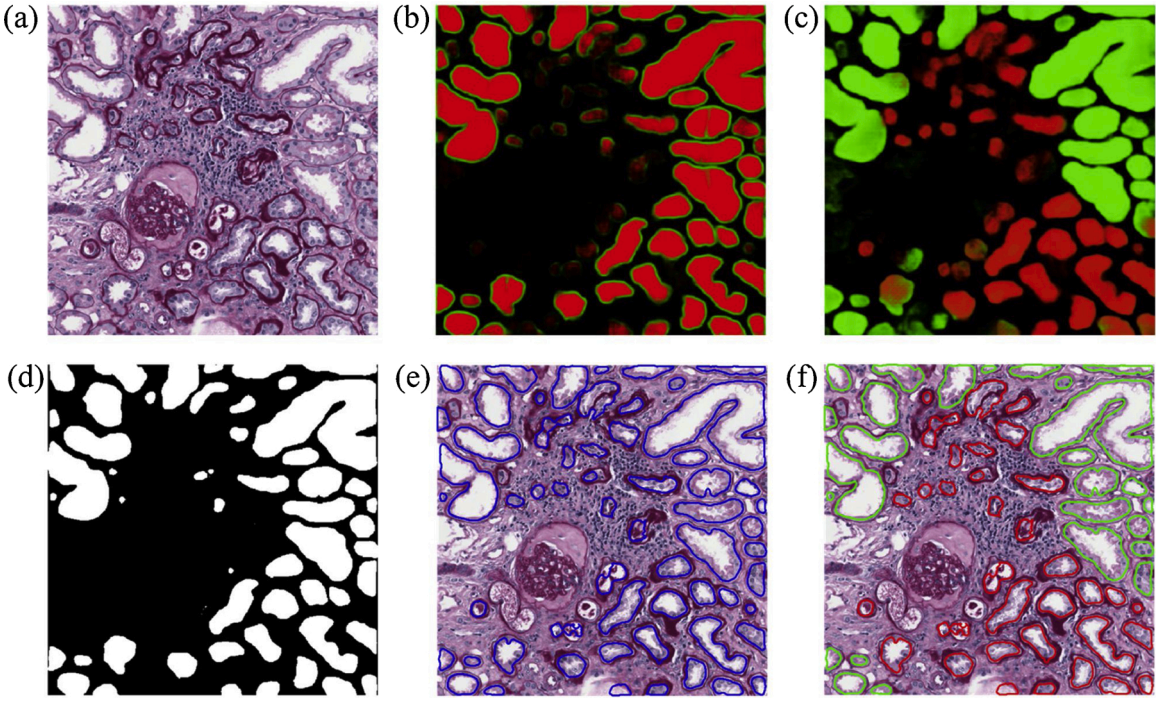

Fig. 6. Tubule segmentation and classification. (a) Original image, (b) softmax of the tubule segmentation network (red: interior of the tubule, green: edge of the tubule), (c) softmax of the tubule classification network (red: atrophic tubule, green: normal tubule), (d) raw tubule detection, (e) final tubule segmentation, (f) kidney tubule classification. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In order to retain the boundary information during the test phase, we synthesize a mirror border of 256 pixels in each direction as shown in Fig. 4. The resulting image, called *extended image*, has a dimension of 1024 × 1024 pixels. A sliding window operator with a size of 512 × 512 is then passed over the extended image with an overlap of 50 % (256 pixels) between consecutive windows. The deep network is applied to each 512 × 512 window, and only the center of each prediction is kept for the creation of the initial softmax. This operation yields a probability map, in which each pixel is assigned a probability of being a healthy glomerulus or a sclerotic one. The resulting softmax is further center cropped to obtain the final softmax with the same size as the input image. The final softmax is treated as an RGB image, where the first layer contains the softmax of sclerotic glomeruli (red) while the second layer (green) includes the softmax of healthy glomeruli.

3.3.2. Post-processing

Once the probability map has been obtained from the deep network (Fig. 5a), the RENTAG algorithm applies a fixed threshold of 0.35 to obtain a raw segmentation of the glomeruli contours. We choose a threshold value lower than 0.5 to increase the sensitivity of the proposed approach. Then, a morphological opening with a disk of 5-pixel radius (equal to 4.67 μm) is carried out to obtain smoother contours. Finally, all the detected shapes with an area less than 700 μ m^2 are deleted as they are too small to be potentially considered as glomeruli (Fig. 5b).

The RENTAG algorithm also detects the basic cellular component within the image: lumen and nuclei. The normalized PAS image is converted to grayscale, Wiener filtered, and thresholded using a fixed value of 245/255 (equal to 96 % of the image maximum) to segment all the white regions. All detected lumen regions are erased to process only PAS stained structures. In order to detect nuclear structures, stain separation (Salvi et al., 2020b) is employed to enhance all the nuclear components. Then, an object-based thresholding is applied to segment the cell nuclei. For each possible threshold point $T \in [0-255]$, the following energy function is computed:

$$E(T) = p_0^2 * var_0 * log(var_0) + p_1^2 * var_1 * log(var_1) \quad (3)$$

where p_0 represents the probability of having gray values equal or lower than T, p_1 is defined as $1 - p_0$, while var_0 and var_1 are the variances of the probability functions of the two classes p_0 and p_1. The optimal thresholding point is then found by minimizing the energy function E in Eq. (3). The result of lumen and nuclei segmentation is shown in Fig. 5c. All the detected structures not containing cell nuclei are deleted from the processing.

The RENTAG algorithm employs the energy model of Chan and Vese (2001) to better delineate the glomeruli boundaries. As can be seen in Fig. 5a-b, the UNET softmax has a higher contrast between glomeruli and background compared to the original RGB image. For this reason, the active contour model is applied directly to the UNET softmax using the raw glomeruli mask (Fig. 5b) as initial contour of the curve. After the level-set, our method applies a structural cleaning step on all the detected regions: all the shapes that do not contain nuclei and/or with an eccentricity lower than 75 % are deleted as they represent tissue artifacts. Since the Bowman's space is absent in the presence of glomerulosclerosis, all the detected sclerotic glomeruli with a lumen area (white region) higher than 10 % are deleted. The final result of the RENTAG algorithm is shown in Fig. 5d.

3.4. Tubular atrophy quantification

3.4.1. CNN semantic segmentation

The RENTAG algorithm adopts the same network architecture described in Section 3.3.1 to detect tubular structures (Fig. 6a). In particular, two different UNET networks are employed to perform tubule segmentation. Both networks are trained using Adam optimization algorithm (Kingma and Ba, 2014) and categorical cross-entropy as loss function. The first CNN, named TSC (tubule segmentation channel), is designed to distinguish foreground pixels (tubule) from background pixels. To train the network, the image pixels are manually labeled into three different classes: interior of the tubule, edge of the tubule or background. The same strategy illustrated in Fig. 4 is adopted to build the network softmax (Fig. 6b). The second CNN, called TCC (tubule

Table 2
Comparison between the proposed algorithm and current state-of-art methods in the glomeruli segmentation (pixel-based metrics). The best results are highlighted in bold.

Method	Subset	Computational Time (sec)	Precision	Recall	DSC
Gallego et al.	TRAIN	24.12 ± 1.63	0.8205 ± 0.2008	0.6939 ± 0.2017	0.7284 ± 0.1824
	TEST	23.92 ± 1.54	0.8257 ± 0.1698	0.7096 ± 0.1912	0.7404 ± 0.1587
Kawazoe et al.	TRAIN	**0.58 ± 0.12**	0.6878 ± 0.1006	0.9066 ± 0.0809	0.7744 ± 0.0697
	TEST	**0.54 ± 0.09**	0.7114 ± 0.1054	0.8916 ± 0.0869	0.7831 ± 0.0685
Kannan et al.	TRAIN	17.18 ± 1.14	0.8898 ± 0.1303	0.7149 ± 0.1995	0.7711 ± 0.1254
	TEST	17.83 ± 1.08	0.8904 ± 0.1278	0.6919 ± 0.1823	0.7545 ± 0.1403
Bueno et al.	TRAIN	1.15 ± 0.17	0.9749 ± 0.0259	0.9321 ± 0.0925	0.9498 ± 0.0667
	TEST	1.24 ± 0.13	0.9674 ± 0.0366	0.9396 ± 0.0486	0.9522 ± 0.0307
Altini et al.	TRAIN	1.18 ± 0.12	0.8786 ± 0.0825	**0.9940 ± 0.0236**	0.9302 ± 0.0533
	TEST	1.21 ± 0.09	0.8936 ± 0.0671	0.9368 ± 0.0793	0.9118 ± 0.0575
CNN no pre-processing[1]	TRAIN	2.24 ± 0.15	**0.9873 ± 0.0266**	0.8799 ± 0.1235	0.9246 ± 0.0876
	TEST	2.26 ± 0.11	**0.9851 ± 0.0188**	0.8758 ± 0.1060	0.9233 ± 0.0675
CNN no post-processing[2]	TRAIN	4.54 ± 0.68	0.9736 ± 0.0331	0.9435 ± 0.0735	0.9563 ± 0.0515
	TEST	4.35 ± 0.72	0.9672 ± 0.0400	0.9296 ± 0.0927	0.9445 ± 0.0689
RENTAG algorithm	TRAIN	6.15 ± 1.03	0.9600 ± 0.0417	0.9634 ± 0.0728	**0.9592 ± 0.0545**
	TEST	5.89 ± 1.86	0.9562 ± 0.0454	**0.9535 ± 0.0649**	**0.9529 ± 0.0433**

[1] CNN with the same architecture shown in Fig. 3 but trained on original images. [2] CNN trained on normalized images but without our post-processing (Section 3.3.2).

classification channel), is designed to semantically segment and classify kidney tubules. In this case, the image pixels are labeled into three different classes depending on whether they belong to a normal tubule, an atrophic tubule or other parts of the tissue (Fig. 6c).

3.4.2. Post-processing

The RENTAG algorithm applies a fixed thresholding of 0.35 on the TSC softmax (Fig. 6b) to obtain a raw segmentation of the inner and border regions of kidney tubules. To refine the segmentation, a custom active contour model (Salvi et al., 2020c) is applied to TSC softmax using the two obtained masks as initial conditions:

$$F = \alpha \cdot |\Gamma| + \beta \int_{\Gamma} (xdy - ydx) + \lambda_1 \cdot \int_{\Omega_{in}} |I(x,y) - \mu_{in}|^2 + \lambda_2 \cdot \int_{\Omega_{out}} |I(x,y) - \mu_{out}|^2 \quad (4)$$

where Γ denotes the curve; μ_{in} and μ_{out} are the mean intensity of the pixels inside and outside the curve; Ω_{in} and Ω_{out} represent the regions inside and outside the curve. Then, the border region mask is subtracted to the inner region mask to obtain a first separation of kidney tubules. Finally, the tubule mask is dilated using a disk element with a radius equal to the thickness of the border mask to retrieve the boundary information (Fig. 6d).

As can be seen from Fig. 6d, most of the tubules in the center of the image are lost as they are too small to be detected by the TSC network. To overcome this limitation, the TCC network is also employed to obtain the final segmentation of the tubules. Fixed thresholding (0.35) is applied to the TCC softmax to obtain a rough contour of normal and atrophic tubules. Then, our strategy applies a cross-check: if a tubule has been segmented by the TCC network and is not present in the tubule mask (Fig. 6d), the shape is added to the final tubule mask. This procedure is repeated for all regions detected by the TCC network. The final segmentation mask is shown in Fig. 6e. The reason for choosing both edge and object detection is to define the spatial limit of each tubule based on the information on the contour (TSC network) and location (TCC network) of each object. In fact, the TSC network allows to separate tubules very close to each other (*contour information*) while the TCC network is able to segment very small tubules (*region information*). Finally, a label (i.e., atrophic or normal) is assigned to each tubule based on the mean value of the TCC softmax (Fig. 6f).

3.5. Performance metrics

Automatic masks are compared with manual ones to evaluate the performance of our strategy in the segmentation of glomeruli and tubules. Several pixel-based metrics as precision, recall, and dice coefficient (DSC) are calculated. Precision is a common metric to assess the false detection of ghost shapes, recall assesses the missed detection of ground truth objects, and DSC measures the spatial overlap between two binary shapes (Zou et al., 2004).

Since the pathologist counts the number of healthy and sclerotic glomeruli to evaluate glomerulosclerosis, two additional object-based metrics are implemented: sensitivity and PPV. These indicators allow you to evaluate how many glomeruli are missing (sensitivity) and how many are false positives (PPV):

$$sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$PPV = \frac{TP}{TP + FP} \quad (6)$$

In the presence of tubular atrophy, the pathologist estimates the percentage of atrophic tubules respect to all tubules within the biopsy. To assess the performance of the proposed strategy for tubular atrophy, two absolute errors (AE) are evaluated: AE_{NUMBER} e AE_{AREA} which are defined as follows:

$$AE_{NUMBER} = \left| \left(\frac{NUM_{ATRO}}{NUM_{TOTAL}} \right)_{MANUAL} - \left(\frac{NUM_{ATRO}}{NUM_{TOTAL}} \right)_{RENTAG} \right| \quad (7)$$

$$AE_{AREA} = \left| \left(\frac{AREA_{ATRO}}{AREA_{TOTAL}} \right)_{MANUAL} - \left(\frac{AREA_{ATRO}}{AREA_{TOTAL}} \right)_{RENTAG} \right| \quad (8)$$

where NUM_{ATRO} represents the number of atrophic tubules, NUM_{TOTAL} is the total number of tubules within the image, $AREA_{ATRO}$ denotes the area of atrophic tubules while $AREA_{TOTAL}$ is the area of all kidney tubules within the image.

4. Results

The results provided by the RENTAG method on the detection of glomeruli are compared both with manual annotations and with previously published works (Gallego et al., 2018; Kawazoe et al., 2018; Kannan et al., 2019; Bueno et al., 2020; Altini et al., 2020). Since the source codes of these works are not publicly available, we have reproduced the pipeline described in each 'Materials and Methods' section for all the compared methods (see Supplementary Material). Specifically, a quantitative comparison is carried out by evaluating both pixel-based metrics (precision, recall and DSC) and object-based metrics

Table 3
Comparison between the proposed algorithm and current state-of-art methods in the glomeruli segmentation (object-based metrics). The best results are highlighted in bold.

Method	Subset	$Sensitivity_{HEALTHY}$	$Sensitivity_{SCLEROTIC}$	$PPV_{HEALTHY}$	$PPV_{SCLEROTIC}$
Gallego et al.	TRAIN	97.24 %	84.74 %	98.72 %	86.30 %
	TEST	94.11 %	15.38 %	60.00 %	50.00 %
Kawazoe et al.	TRAIN	93.28 %	79.66 %	99.20 %	87.03 %
	TEST	92.07 %	69.23 %	96.87 %	90.00 %
Kannan et al.	TRAIN	87.21 %	61.01 %	96.93 %	62.06 %
	TEST	91.17 %	84.61 %	96.87 %	61.11 %
Bueno et al.	TRAIN	97.24 %	84.74 %	98.72 %	86.20 %
	TEST	97.02 %	100 %	98.98 %	86.66 %
Altini et al.	TRAIN	99.25 %	100 %	99.00 %	**100 %**
	TEST	95.04 %	84.61 %	100 %	73.73 %
CNN no pre-processing[1]	TRAIN	95.53 %	88.13 %	100 %	69.33 %
	TEST	91.17 %	92.30 %	100 %	50.00 %
CNN no post-processing[2]	TRAIN	98.49 %	**100 %**	**100 %**	76.62 %
	TEST	96.07 %	100 %	100 %	86.66 %
RENTAG algorithm	TRAIN	**99.49 %**	98.30 %	99.74 %	96.66 %
	TEST	**97.02 %**	**100 %**	**100 %**	**92.85 %**

[1] CNN with the same architecture shown in Fig. 3 but trained on original images. [2] CNN trained on normalized images but without our post-processing (Section 3.3.2).

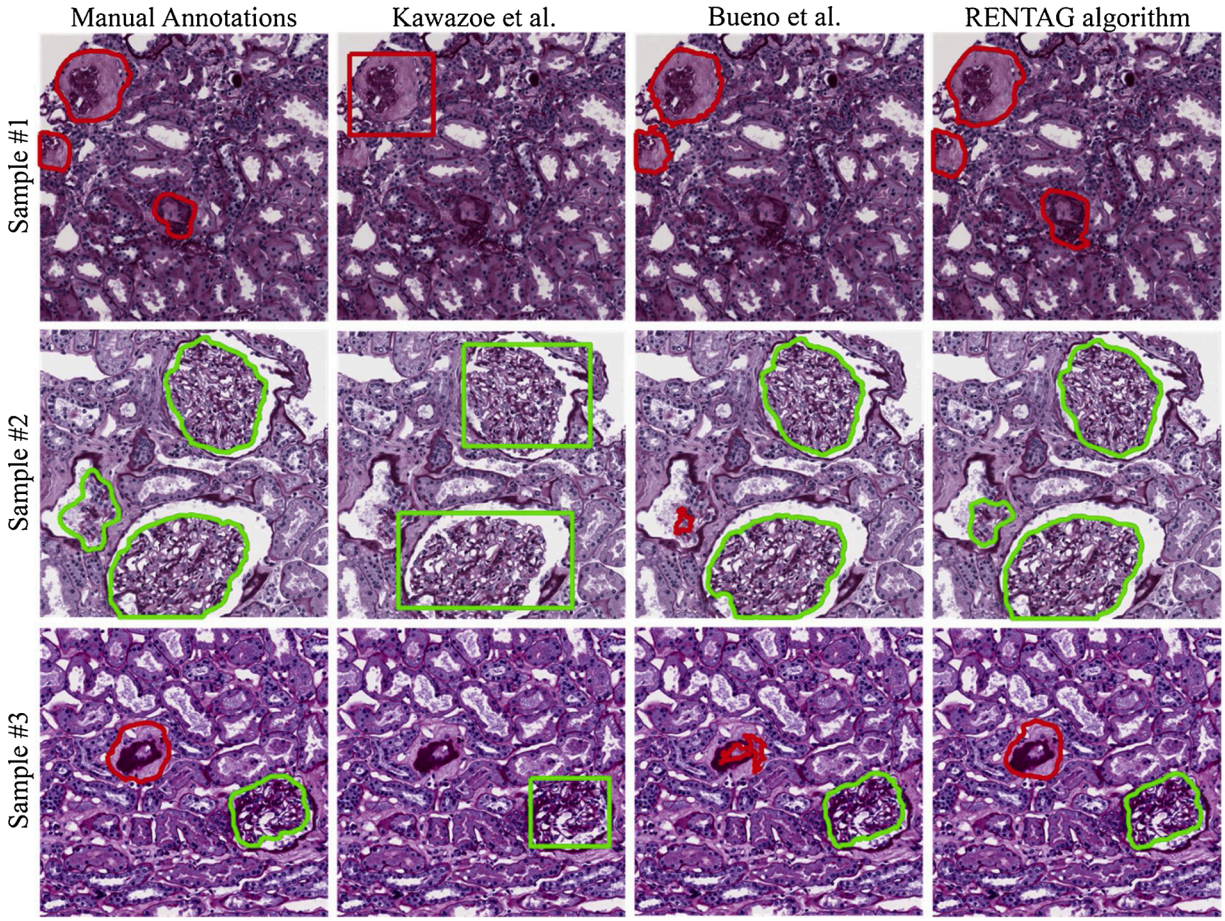

Fig. 7. Visual performance between RENTAG algorithm and two of the best published deep learning methods for glomeruli detection (red: atrophic tubule, green: normal tubule). Three different samples are illustrated in rows while segmentation results are shown in columns. The original image along with manual annotations is displayed in the first column. Compared methods are shown in the second and third column, while the results of our strategy is presented in the last column. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(sensitivity and PPV). To demonstrate the superiority of our strategy, we also evaluate the results obtained using the three-class UNET without the stain normalization (CNN no pre-processing) and the three-class UNET without the proposed post-processing (CNN no post-processing). The processing is performed on a workstation with a 4.1 GHz octa-core CPU with 32 Gb of RAM. Tables 2 and 3 show the results obtained in segmenting glomerular structures.

For all the pixel-based metrics (precision, recall and DSC), our method outperforms the state-of-art techniques in both TRAIN and TEST set. The RENTAG algorithm also gets the best object-based metrics

Table 4
Performance of the RENTAG algorithm in kidney tubule segmentation.The best results are highlighted in bold.

Method	Subset	Computational Time (sec)	Precision	Recall	DSC
TSC network	TRAIN	**8.18 ± 3.82**	0.9157 ± 0.0230	0.8211 ± 0.0429	0.8653 ± 0.0280
	TEST	**8.41 ± 3.97**	0.9145 ± 0.0180	0.8162 ± 0.0171	0.8584 ± 0.0125
TCC network	TRAIN	9.38 ± 3.86	0.9178 ± 0.0298	0.8143 ± 0.0334	0.8629 ± 0.0263
	TEST	9.54 ± 3.92	0.9104 ± 0.0283	0.8206 ± 0.0189	0.8624 ± 0.0199
RENTAG algorithm	TRAIN	15.32 ± 2.49	**0.9504 ± 0.0258**	**0.8889 ± 0.0367**	**0.9182 ± 0.0260**
	TEST	14.63 ± 2.70	**0.9455 ± 0.0272**	**0.8915 ± 0.0212**	**0.9174 ± 0.0165**

Table 5
Absolute errors in the assessment of tubular atrophy for both TRAIN and TEST set.

Method	Subset	AE_{NUMBER}	AE_{AREA}
RENTAG algorithm	TRAIN	1.95 % ± 2.53 %	1.56 % ± 2.63 %
	TEST	2.43 % ± 3.13 %	1.96 % ± 2.54 %

(sensitivity and PPV) for healthy and sclerotic glomeruli. In the TEST set, our strategy does not miss any sclerotic glomerulus ($sensitivity_{SCLEROTIC}$ = 100 %) and does not mistakenly segment any healthy glomerulus ($PPV_{HEALHTY}$ = 100 %). More interestingly, the proposed post-processing allows to increase the PPV of sclerotic glomeruli up to 20 % (CNN no post-processing vs. RENTAG) with only 1.61 s of additional processing. Overall, the high average metrics coupled with a low standard deviation demonstrate the robustness of the proposed algorithm. Fig. 7 shows a visual comparison between the RENTAG algorithm and previously published works. Our approach is able to correctly segment the glomeruli found on the edges of the image (Fig. 7 – Sample #1). The combination of semantic segmentation and active contours allows to obtain a more accurate glomerular profile respect to other methods (Fig. 7 – Sample #2 and #3).

To evaluate the performance of the RENTAG algorithm in tubule segmentation, the same pixel-based metrics adopted for glomerulosclerosis are used (precision, recall, DSC). In particular, we compared the performance of the individual networks (TSC: Tubule Segmentation Channel, TCC: Tubule Classification Channel) and the result provided by the combination of the two networks (RENTAG algorithm). Table 4 summarize the performance of our strategy for both TRAIN and TEST set. As can be seen, the RENTAG algorithm outperforms the individual networks. This result demonstrate that this joint effort (TSC + TCC network) performs better than the single approach. Our method exhibits excellent performance in tubule segmentation, obtaining a precision higher than 95 %. In addition, the DSC is stable for both sets (91.82 % on TRAIN and 91.74 % on TEST), thus demonstrating the solidity of our algorithm. Moreover, the absolute error in the evaluation of tubular atrophy (AE_{NUMBER} and AE_{AREA}) is negligible for each image of the dataset (Table 5).

The segmentation result is illustrated in Fig. 8, where images with different tubules morphology and spatial density are presented as explanatory examples. Our algorithm is able to correctly locate and classify each tubule within the image.

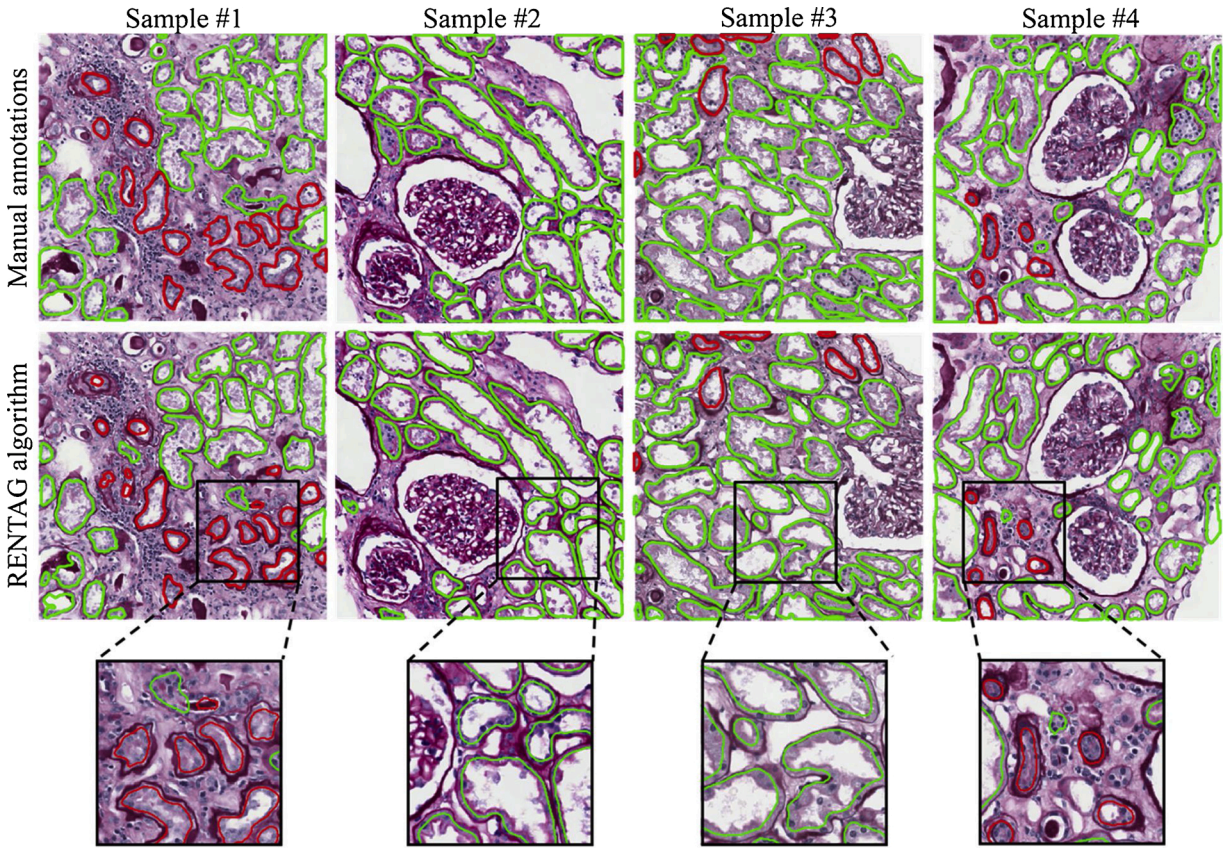

Fig. 8. Tubule segmentation and classification for different samples with variation of stain intensity, density and tubules morphology (red: atrophic tubule, green: normal tubule). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

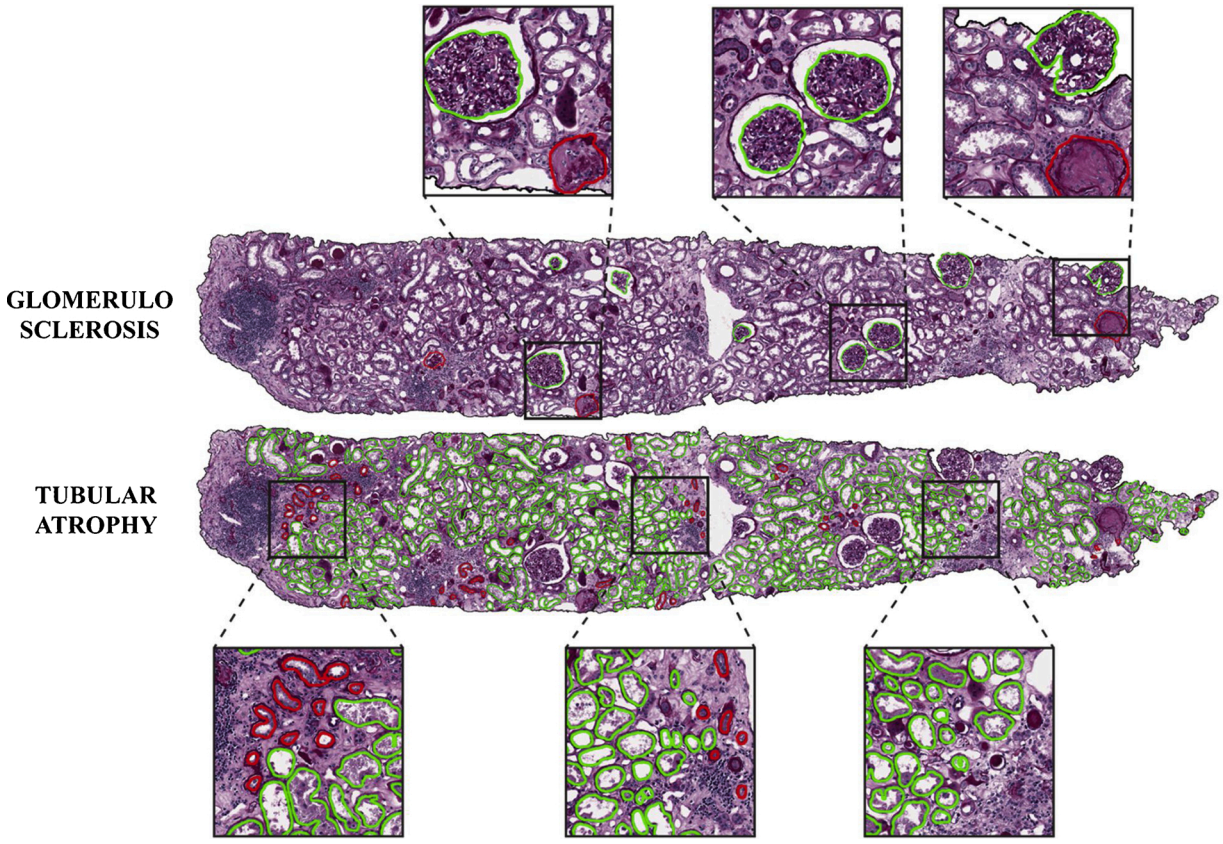

Fig. 9. Result of RENTAG processing on a whole-slide image. Glomerulosclerosis evaluation: healthy and sclerotic glomeruli are highlighted in green and red respectively. During the assessment of tubular atrophy, atrophic tubules are shown in red while normal tubules are represented in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Since the pathologist assesses the glomerulosclerosis and tubular atrophy on whole-slide images (WSIs), we extend our approach on entire biopsies using a sliding-windows approach. An expert pathologist takes at least 15 min to assess the degree of glomerulosclerosis and tubular atrophy while our algorithm is able to process the entire biopsy in less than 3 min. Fig. 9 shows the application of the RENTAG algorithm on a WSI. The use of automated algorithms can provide a more consistent and accurate case assessment, ensuring quality controls and reducing the need for a large number of pathologists.

5. Discussion and conclusions

Pathologists have to categorize and summarize their findings in a schematic but complete "language." Therefore, grading and scoring systems have always been welcome and thoroughly developed to standardize the reports and propose a clinically applicable code. However, inter- and intra-pathologist evaluation dissimilarity cause significant limitations, even in the most severe score system. Transplant pathology makes no exception. During this relatively recent pathology field development, the need for a worldwide common way of defining transplant-related pathology findings becomes urgent and crucial. In kidney transplantation, this need led to the Karpinski score, a four-grade morphology-based score to evaluate and simultaneously synthesize the injuries revealed during pre-transplantation biopsies analysis. Since time is crucial in transplant applications, quick and accurate diagnoses are required from pathologists. However, this procedure is not free from limitations, also because of reproducibility and timing feasibility.

In this study, we present a fully automated method for quantitative assessment of glomerulosclerosis and tubular atrophy in kidney histopathological images. Detection of kidney structures is a real challenge due to the variability of shape and size, especially in pathological conditions. Thanks to the stain normalization step, our strategy is capable of automatically detecting glomeruli and tubules in images with different staining intensity. The proposed method is developed and validated on 830 PAS stained images of kidney tissue and results are compared with manual annotations of an expert pathologist. During glomerulosclerosis assessment, the RENTAG algorithm achieves the highest precision, recall, and DSC compared to five published methods in both TRAIN and TEST set. More importantly, our approach outperforms the compared algorithms for all the object-based metrics (Table 3). Our method also achieves satisfactory performance in tubule segmentation and classification, obtaining a DSC over 91 %. As can be seen from Fig. 8, the RENTAG algorithm maintains high performance even in the presence of different tubular morphology and density, thus demonstrating the robustness of the proposed approach.

These high performances are mainly due to the combination of semantic segmentation with an ad-hoc post-processing. By detecting cell nuclei and lumen regions, the RENTAG algorithm is able to delete almost all the false-positive shapes segmented by the CNN. Level set and morphological cleaning are then employed to refine the segmentation result. To quantify tubular atrophy, a novel approach that combines two different deep networks (TSC and TCC) is employed. The joint effort of both networks allows to generate an instance segmentation of each tubule. As far as we know, no automated solution has been proposed so far for segmentation and classification of renal tubules. Using the strategy shown in Fig. 4, the RENTAG algorithm is also able to preserve the boundary information during the inference phase. This particular strategy allows to accurately segment objects on the edges of the image. The main limitation of the RENTAG algorithm refers to the tubule segmentation. As can be observed from Table 4 and Fig. 8, our method slightly underestimates the tubule boundaries (recall equal to 0.88). In addition, images need to be acquired at 10x or greater magnification.

Using a lower resolution, the cellular structures segmentation may fail due to poor quality images.

In this study, we demonstrated that the analysis of preimplantation biopsies with a fully automated algorithm could practically support the pathologists' assessment of the Karpinski score. In the near future, kidney transplantation would probably be enriched with an increasing number of ECDs kidneys, posing more significant challenges to the pathology diagnostics activity. In this setting, the introduction of a fully automated deep learning-based algorithm in the daily routine could represent a crucial turning point for diagnostics improvement in pre-transplant kidney panorama. In the future, we will integrate the assessment of arteriolar narrowing and interstitial fibrosis (Salvi et al., 2020b) within the RENTAG algorithm in order to create the first automated Karpinski scoring system.

CRediT authorship contribution statement

Massimo Salvi: Methodology, Software, Writing – Original Draft. **Alessandro Mogetta**: Software, Writing – Review & Editing. **Alessandro Gambella**: Curation, Writing – Review & Editing. **Luca Molinaro**: Investigation, Curation, Resources. **Antonella Barreca**: Writing – Review & Editing. **Mauro Papotti**: Supervision. **Filippo Molinari**: Conceptualization, Supervision.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

The authors would like to acknowledge all the laboratory technicians of the Division of Pathology (Department of Oncology, Turin, Italy) for their help in digitizing histological.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.compmedimag.2021.101930.

References

Altini, N., Cascarano, G.D., Brunetti, A., Marino, F., Rocchetti, M.T., Matino, S., Venere, U., Rossini, M., Pesce, F., Gesualdo, L., 2020. Semantic segmentation framework for glomeruli detection and classification in kidney histological sections. Electronics 9, 503.

Anghel, A., Stanisavljevic, M., Andani, S., Papandreou, N., Rüschoff, J.H., Wild, P., Gabrani, M., Pozidis, H., 2019. A high-performance system for robust stain normalization of whole-slide images in histopathology. Front. Med. 6.

Belhomme, P., Toralba, S., Plancoulaine, B., Oger, M., Gurcan, M.N., Bor-Angelier, C., 2015. Heterogeneity assessment of histological tissue sections in whole slide images. Comput. Med. Imaging Graph. 42, 51–55.

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

Bukowy, J.D., Dayton, A., Cloutier, D., Manis, A.D., Staruschenko, A., Lombard, J.H., Woods, L.C.S., Beard, D.A., Cowley, A.W., 2018. Region-based convolutional neural nets for localization of glomeruli in trichrome-stained whole kidney sections. J. Am. Soc. Nephrol. 29, 2081–2088.

Chan, T.F., Vese, L.A., 2001. Active contours without edges. IEEE Trans. Image Process. 10, 266–277.

Chen, C., Huang, Y., Fang, P., Liang, C., Chang, R., 2020. A computer-aided diagnosis system for differentiation and delineation of malignant regions on whole-slide prostate histopathology image using spatial statistics and multidimensional DenseNet. Med. Phys. 47, 1021–1033.

Gadermayr, M., Dombrowski, A.-K., Klinkhammer, B.M., Boor, P., Merhof, D., 2019. CNN cascades for segmenting sparse objects in gigapixel whole slide images. Comput. Med. Imaging Graph. 71, 40–48.

Gallego, J., Pedraza, A., Lopez, S., Steiner, G., Gonzalez, L., Laurinavicius, A., Bueno, G., 2018. Glomerulus classification and detection based on convolutional neural networks. J. Imaging 4, 20.

Hassan, A., Halawa, A., 2015. Dual kidney transplant. Exp. Clin. Transplant. 13, 500–509.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 770–778.

Janowczyk, A., Madabhushi, A., 2016. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. J. Pathol. Inform. 7.

Jefferson, J.A., Shankland, S.J., 2014. The pathogenesis of focal segmental glomerulosclerosis. Adv. Chronic Kidney Dis. 21, 408–416.

Jha, K.K., Dutta, H.S., 2019. Mutual Information based hybrid model and deep learning for Acute Lymphocytic Leukemia detection in single cell blood smear images. Comput. Methods Programs Biomed. 179, 104987.

Kannan, S., Morgan, L.A., Liang, B., Cheung, M.G., Lin, C.Q., Mun, D., Nader, R.G., Belghasem, M.E., Henderson, J.M., Francis, J.M., 2019. Segmentation of glomeruli within trichrome images using deep learning. Kidney Int. reports 4, 955–962.

Karpinski, J., Lajoie, G., Cattran, D., Fenton, S., Zaltzman, J., Cardella, C., Cole, E., 1999. Outcome of kidney transplantation from high-risk donors is determined by both structure and function. Transplantation 67, 1162–1167.

Kawazoe, Y., Shimamoto, K., Yamaguchi, R., Shintani-Domoto, Y., Uozaki, H., Fukayama, M., Ohe, K., 2018. Faster r-cnn-based glomerular detection in multistained human whole slide images. J. Imaging 4, 91.

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv Prepr. arXiv1412.6980.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, pp. 1097–1105.

Litjens, G., Sánchez, C.I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen-Van De Kaa, C., Bult, P., Van Ginneken, B., Van Der Laak, J., 2016. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci. Rep. 6, 26286.

Mazzucco, G., Magnani, C., Fortunato, M., Todesco, A., Monga, G., 2010. The reliability of pre-transplant donor renal biopsies (PTDB) in predicting the kidney state. A comparative single-centre study on 154 untransplanted kidneys. Nephrol. Dial. Transplant. 25, 3401–3408.

Metzger, R.A., Delmonico, F.L., Feng, S., Port, F.K., Wynn, J.J., Merion, R.M., 2003. Expanded criteria donors for kidney transplantation. Am. J. Transplant. 3, 114–125.

Port, F.K., Bragg-Gresham, J.L., Metzger, R.A., Dykstra, D.M., Gillespie, B.W., Young, E. W., Delmonico, F.L., Wynn, J.J., Merion, R.M., Wolfe, R.A., 2002. Donor characteristics associated with reduced graft survival: an approach to expanding the pool of kidney donors1. Transplantation 74, 1281–1286.

Remuzzi, G., Perico, N., 1998. Protecting single-kidney allografts from long-term functional deterioration. J. Am. Soc. Nephrol. 9, 1321–1332.

Rosengard, B.R., Feng, S., Alfrey, E.J., Zaroff, J.G., Emond, J.C., Henry, M.L., Garrity, E. R., Roberts, J.P., Wynn, J.J., Metzger, R.A., 2002. Report of the Crystal City meeting to maximize the use of organs recovered from the cadaver donor. Am. J. Transplant. 2, 701–711.

Salvi, M., Acharya, U.R., Molinari, F., Meiburger, K.M., 2020a. The impact of pre-and post-image processing techniques on deep learning frameworks: a comprehensive review for digital pathology image analysis. Comput. Biol. Med. 128, 104129.

Salvi, M., Mogetta, A., Meiburger, K.M., Gambella, A., Molinaro, L., Barreca, A., Papotti, M., Molinari, F., 2020b. Karpinski score under digital investigation: A fully automated segmentation algorithm to identify vascular and stromal injury of Donors' kidneys. Electron 9 (10). https://doi.org/10.3390/electronics9101644.

Salvi, M., Molinaro, L., Metovic, J., Patrono, D., Romagnoli, R., Papotti, M., Molinari, F., 2020c. Fully automated quantitative assessment of hepatic steatosis in liver transplants. Comput. Biol. Med. 123, 103836.

Sharma, H., Zerbe, N., Klempert, I., Hellwich, O., Hufnagl, P., 2017. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. Comput. Med. Imaging Graph. 61, 2–13.

Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J.-M., Ciompi, F., van der Laak, J., 2019. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Med. Image Anal. 58, 101544.

Wolfe, R.A., Ashby, V.B., Milford, E.L., Ojo, A.O., Ettenger, R.E., Agodoa, L.Y.C., Held, P. J., Port, F.K., 1999. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. N. Engl. J. Med. 341, 1725–1730.

Yan, R., Ren, F., Wang, Z., Wang, L., Zhang, T., Liu, Y., Rao, X., Zheng, C., Zhang, F., 2020. Breast cancer histopathological image classification using a hybrid deep neural network. Methods 173, 52–60.

Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M.C., Kaus, M.R., Haker, S.J., Wells III, W.M., Jolesz, F.A., Kikinis, R., 2004. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. Acad. Radiol. 11, 178–189.