



ELSEVIER

Contents lists available at ScienceDirect

Data in brief

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



Data Article

Data for glomeruli characterization in histopathological images



Gloria Bueno ^{a,*}, Lucia Gonzalez-Lopez ^{b,1},
Marcial Garcia-Rojo ^{c,1}, Arvydas Laurinavicius ^{d,1},
Oscar Deniz ^{a,1}

^a VISILAB, Universidad de Castilla-La Mancha, Ciudad Real, Spain

^b Pathology Department, Hospital Universitario de Ciudad Real, Ciudad Real, Spain

^c Pathology Department, Hospital Universitario Puerta del Mar, Cádiz, Spain

^d Faculty of Medicine and National Center of Pathology, Vilnius, Lithuania

ARTICLE INFO

Article history:

Received 7 February 2020

Accepted 13 February 2020

Available online 24 February 2020

Keywords:

Glomeruli identification

Normal glomerulus

Global sclerotic glomerulus

Whole slide image

Digital pathology

ABSTRACT

The data presented in this article is part of the whole slide imaging (WSI) datasets generated in European project AIDPATH ² This data is also related to the research paper entitled "Glomerulosclerosis Identification in Whole Slide Images using Semantic Segmentation", published in Computer Methods and Programs in Biomedicine Journal [1]. In that article, different methods based on deep learning for glomeruli segmentation and their classification into normal and sclerotic glomerulus are presented and discussed. The raw data used is described and provided here. In addition, the detected glomeruli are also provided as individual image files. These data will encourage research on artificial intelligence (AI) methods, create and compare fresh algorithms, and measure their usability in quantitative nephropathology.

© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

DOI of original article: <https://doi.org/10.1016/j.cmpb.2019.105273>.

* Corresponding author.

E-mail address: gloria.bueno@uclm.es (G. Bueno).

¹ All authors contributed equally to this work.

² <http://aidpath.eu>

<https://doi.org/10.1016/j.dib.2020.105314>

2352-3409/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject	<i>Pathology and Medical Technology</i>
Specific subject area	<i>Digital images in pathology. Annotated and classify whole slide image (WSI). Histopathology.</i>
Type of data	<i>Images, Whole Slide Images.</i>
How data were acquired	<i>Hardware: Whole Slide Digital Image Scanner (Leica Aperio ScanScope CS scanner) and extracted into an SVS file format. Software for annotation: Aperio ImageScope. Software for detection and extraction: VISILAB group software described in Ref. [1].</i>
Data format	<i>Raw data: a) original images in SVS format Classified: Detected glomeruli to be used for classification in PNG format.</i>
Parameters for data collection	<i>Tissue samples were collected with a biopsy needle having an outer diameter between 100 μm and 300 μm. Afterwards, paraffin blocks were prepared using tissue sections of 4 μm and stained using Periodic acid–Schiff (PAS). Then, images at 20x magnification were selected.</i>
Description of data collection	<i>The tissue samples were scanned at 20x with a Leica Aperio ScanScope CS scanner.</i>
Data source location	<i>VISILAB, Universidad de Castilla-La Mancha, Ciudad Real, Spain</i>
Data accessibility	<i>Hosted by Mendeley at: https://data.mendeley.com/datasets/k7nvtgn2x6/3</i>
Related research article	<i>Author's name: Gloria Bueno, M. Milagro Fernandez-Carrobles, Lucia Gonzalez-Lopez, Oscar Deniz Title: Glomerulosclerosis Identification in Whole Slide Images using Semantic Segmentation Journal: Computer Methods and Programs in Biomedicine https://doi.org/10.1016/j.cmpb.2019.105273</i>

Value of the Data

- These data can be used for benchmarking to encourage further research on AI methods applied to digital pathology in nephrology.
- The additional value of this data is that it has been acquired and evaluated by expert pathologists from different European countries.
- All researches in digital pathology can benefit from these data, to test classification algorithms. And particularly for glomeruli identification in nephrology studies.
- This data can be used for further development and new experiments in glomeruli classification with more classes, like focal glomeruli besides normal and sclerotic glomeruli.
- The data will assist to provide further insights for nephropathologists, allowing to create novel diagnostic tools.

1. Data

The data is composed of two datasets:

1. DATASET_A: Raw data with 31 whole slide images (WSI) in SVS format. The size of the WSI range between 21651×10498 pixels and 49799×32359 pixels acquired at 20x. The images contain different types of glomeruli that were detected using the algorithms explained at [1]. The detected glomeruli are provided in DATASET_B.
2. DATASET_B: 2340 images with a single glomerulous, 1170 normal glomeruli and 1170 sclerosed glomeruli. All of them are in PNG format.

2. Experimental design, materials, and methods

The introduction of digital pathology to nephrology offers a platform to develop fresh methodologies and protocols for morphometric and computer-aided renal biopsies evaluation. Digital imaging application to pathology may provide significant advance for clinical studies and translational research. The international application of this new technology is driving new collaborative methods, providing

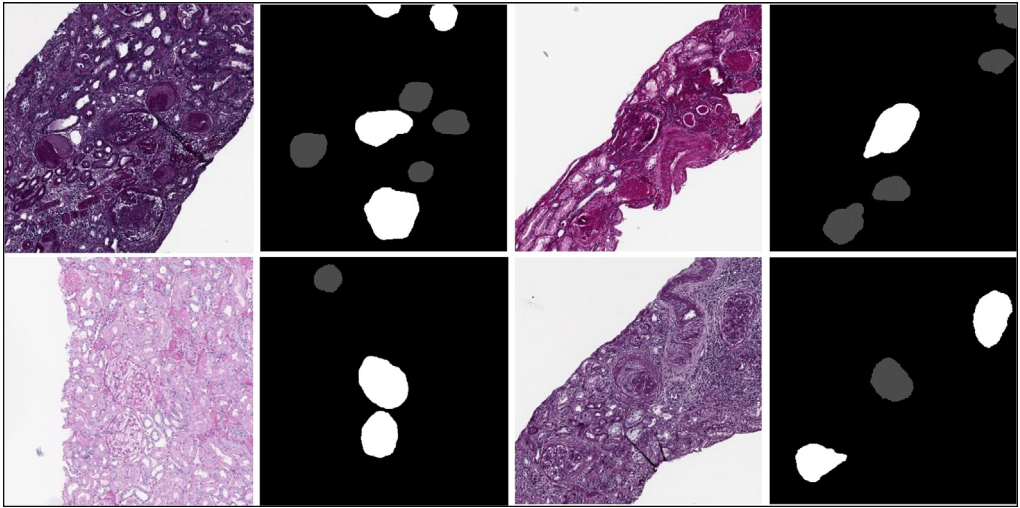


Fig. 1. Samples of the DATASET_A, 1st and 3rd columns correspond to original images (raw data), 2nd and 4th columns are the masks of the segmented regions, that is the ground truth of the three classes considered: non-glomerular structures, normal glomeruli and sclerosed glomeruli. a) normal glomeruli, b) sclerosed glomeruli.

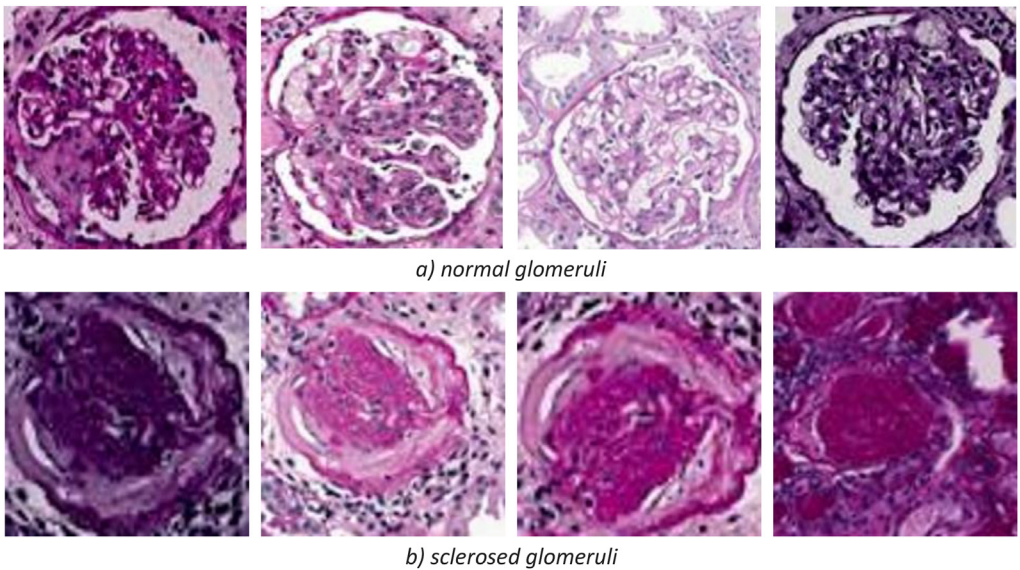


Fig. 2. Samples of the DATASET_B provided for classification of two classes, normal and sclerosed glomeruli.

unique opportunities but also challenges in education [2] and clinical practice [3]. One of the main challenges is the need for data acquired, evaluated and annotated to develop robust algorithms suitable for diagnosis [4].

Objective instruments to quantify and characterize glomeruli are needed for glomerulosclerosis assessment. The number of glomeruli is often used to determine the adequacy of a kidney biopsy and pathologists tend to underreport the number of glomeruli on a biopsy specimen [5].

The European project AIDPATH² aware of the above-mentioned challenges carried out a data set of different whole slide images for breast and kidney tissue acquired by different European institutions under the same protocol. In this paper we supply some of the kidney tissue prepared at three different Pathology Department at Hospital Universitario de Ciudad Real (ES), Hospital Universitario Puerta del Mar (ES) and National Center of Pathology (LT) for glomerulosclerosis analysis.

Tissue samples were collected with a biopsy needle having an outer diameter between 100 μm and 300 μm . Afterwards, paraffin blocks were prepared using tissue sections of 4 μm and stained using PAS. Then, images at 20x magnification were selected. PAS stain is commonly used due to its efficiency dyeing polysaccharides, which are present in kidney tissue and in highlighting glomerular basement membranes [6,7]. A dataset of 47 kidney WSIs was obtained. Images at 20x magnification were selected since this magnification maintain image quality and information at the same as allows to obtain valuable results reducing computational time significantly.

Once WSIs at 20x magnification were collected, they were split into 2000×2000 pixels patches selecting only those which contained tissue. This set of patches was examined and labeled into three classes: i) Non-Glomerular structures: kidney tissue structures such as proximal and distal tubules, blood vessels, connective tissue stroma or inflammatory cells; ii) Normal Glomeruli characterized by thin glomerular capillary loops, a regular number of endothelial and mesangial cells. The aspect of glomerulus surrounding tubules is normal and iii) Sclerosed Glomeruli where the whole (or nearly the whole) glomerulus presents sclerosis.

The output of the previous steps was a dataset with a total of 1055 kidney tissue images, i.e., WSI subsamples with a size of 2000×2000 pixels. Glomeruli contours were annotated by pathologists using the Aperio ImageScope software.³ Then, per each subsample a gray mask was created containing: i) black pixels for the non-glomerular structures, ii) white pixels for the normal glomeruli and iii) gray pixels for the sclerosed glomeruli. As a result, 1245 glomerular structures were annotated, 303 of these were sclerosed glomeruli and 942 were normal glomeruli. These structures were extracted and data augmentation was performed onto these original images applying rotations of 90° and 270° , vertical flip and color transfer with 5 different references taken from the WSI provided at the Hospitals involved (see the associated publication for further details [1]). The purpose of data augmentation was twofold, 1st to augment the data as input of deep learning algorithms and 2nd to balance the number of normal and sclerosed glomeruli.

The raw data used (DATASET_A) together with the detected glomeruli (DATASET_B) are provided alongside this publication. Figs. 1 and 2 illustrate patches of the raw data provided in this article and the detected glomeruli.

Acknowledgments

The authors acknowledge financial support from the AIDPATH² European project num. 612471, <http://aidpath.eu>.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] G. Bueno, M.M. Fernandez-Carrobles, L. Gonzalez-Lopez, O. Deniz, Glomerulosclerosis identification in whole slide images using semantic segmentation, *Comput. Methods Progr. Biomed.* 184 (2019) 105273, <https://doi.org/10.1016/j.cmpb.2019.105273>. Dec 19.
- [2] V. Della Mea, A. Carbone, C. Di Loreto, et al., Teaching digital pathology: the international school of digital pathology and proposed syllabus, *J. Pathol. Inf.* (2017) 1–27, https://doi.org/10.4103/jpi.jpi_17_17.

³ <https://www.leicabiosystems.com/digital-pathology/manage/aperio-imagescope/>.

- [3] L. Barisoni, C. Gimpel, R. Kain, et al., Digital pathology imaging as a novel platform for standardization and globalization of quantitative nephropathology, *Clin. Kidney J.* 10 (2) (2017) 176–187.
- [4] G. Bueno, M.M. Fernández-Carrobles, O. Deniz, M. García-Rojo, New trends of emerging technologies in digital pathology, *Pathobiology* 83 (2016) 61–69.
- [5] R.S. Niznik, C.L. Lopez, W.K. Kremers, et al., Global glomerulosclerosis 421 in kidney biopsies with differing amounts of cortex: a clinical-pathologic 422 correlation study, *Kidney Med.* 1 (4) (2019) 153–161.
- [6] R.A. Risdon, D.R. Turner, *Atlas of Renal Pathology, Current Histopathology*, MTP Press, 1980. Universidad de California, 1980.
- [7] R.R. Robinson, A.Y. Barakat, *Renal Disease in Children: Clinical Evaluation and Diagnosis*, Springer, New York, 2012.