

Received 19 October 2023; revised 12 August 2024; accepted 3 September 2024. Date of publication 10 September 2024;
date of current version 19 September 2024.

Digital Object Identifier 10.1109/JTEHM.2024.3455941

Enhancing Podocyte Degenerative Changes Identification With Pathologist Collaboration: Implications for Improved Diagnosis in Kidney Diseases

GEORGE OLIVEIRA BARROS^{1,8}, JOSÉ NATHAN ANDRADE MULLER DA SILVA²,
HENRIQUE MACHADO DE SOUSA PROENÇA³, STANLEY ALMEIDA ARAÚJO⁴,
DAVID CAMPOS WANDERLEY⁴, LUCIANO REBOUÇAS DE OLIVEIRA⁵,
WASHINGTON LUIS CONRADO DOS-SANTOS⁶, ANGELO AMANCIO DUARTE⁷,
AND FLAVIO DE BARROS VIDAL⁸

¹Instituto Federal Goiano, Goiania 76300-000, Brazil

²Department of Pathology, Federal University of Bahia, Salvador 40110-909, Brazil

³Department of Pathology, Kidney and Hypertension Hospital, Oswaldo Ramos Foundation, São Paulo 04038-002, Brazil

⁴Center for Electron Microscopy, Institute of Nephropathology, Federal University of Minas Gerais, Belo Horizonte 31270-901, Brazil

⁵Intelligent Vision Research Laboratory, Institute of Computing, Federal University of Bahia, Salvador 40170-115, Brazil

⁶Gonçalo Moniz Institute, Oswaldo Cruz Foundation, Salvador 40296-710, Brazil

⁷Department of Technology, State University of Feira de Santana, Feira de Santana 44036-900, Brazil

⁸Department of Computer Science, University of Brasilia, Brasilia 70910-900, Brazil

CORRESPONDING AUTHOR: G. O. BARROS (george.barros@ifgoiano.edu.br)

This work was supported in part by the PathoSpotter Project through the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) under Grant TO-P0008/15, in part by the Universidade Estadual de Feira de Santana under Grant FINAPESQ TO-074/2021, in part by the Inova Fundação Oswaldo Cruz (FIOCRUZ) grant, and in part by the University of Brasilia under Grant DPI/DPG/BCE 01/2024. The work of Washington Luis Conrado dos-Santos and Luciano Rebouças de Oliveira was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under Grant 306779/2017 and Grant 308580/2021-4.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committees of the Gonçalo Moniz Institute, Oswaldo Cruz Foundation (CPqGM/FIOCRUZ), under Approval Nos. 188/09 and 1.817.574.

This article has supplementary downloadable material available at <https://doi.org/10.1109/JTEHM.2024.3455941>, provided by the authors.

ABSTRACT Podocyte degenerative changes are common in various kidney diseases, and their accurate identification is crucial for pathologists to diagnose and treat such conditions. However, this can be a difficult task, and previous attempts to automate the identification of podocytes have not been entirely successful. To address this issue, this study proposes a novel approach that combines pathologists' expertise with an automated classifier to enhance the identification of podocytopathies. The study involved building a new dataset of renal glomeruli images, some with and others without podocyte degenerative changes, and developing a convolutional neural network (CNN) based classifier. The results showed that our automated classifier achieved an impressive 90.9% f-score. When the pathologists used as an auxiliary tool to classify a second set of images, the medical group's average performance increased significantly, from $91.4 \pm 12.5\%$ to $96.1 \pm 2.9\%$ of f-score. Fleiss' kappa agreement among the pathologists also increased from 0.59 to 0.83. Conclusion: These findings suggest that automating this task can bring benefits for pathologists to correctly identify images of glomeruli with podocyte degeneration, leading to improved individual accuracy while raising agreement in diagnosing podocytopathies. This approach could have significant implications for the diagnosis and treatment of kidney diseases.

INDEX TERMS Computational nephropathology, podocyte degenerative changes, glomeruli, deep learning, decision-making.

Clinical impact: The approach presented in this study has the potential to enhance the accuracy of medical diagnoses for detecting podocyte abnormalities in glomeruli, which serve as biomarkers for various glomerular diseases.

I. INTRODUCTION

PODOCYTE lesions can be caused by genetic predisposition or by kidney injury resulting from infection,

toxicity, hemodynamics, or obesity. These lesions may be severe, resulting in podocyte histological changes such as cell swelling, vacuolization, cell detachment of the glomerular

basement membrane, binucleation, and proliferation. Severe podocyte lesions correlate with nephrotic level proteinuria, or nephrotic syndrome [1], and can be present in collapsing nephropathy, focal segmental glomerulosclerosis, lupus nephropathy [2], diabetic nephropathy (DN) [3], [4] and immunoglobulin A nephropathy (IgAN) [5].

Severe podocyte lesions may also induce genetic program in tubular cells that provokes tubulointerstitial inflammation, fibrosis, and tubular atrophy or favors the development of focal and segmental glomerular sclerosis, contributing to progressive loss of renal function [4], [5]. Although reversible slight podocyte lesions can usually be only perceived by electron microscopy analysis, severe podocyte degeneration can be diagnosed through the histological study of renal glomeruli on light microscopy. In recent years, several automated, machine learning-based methods have been proposed to aid in the visual analysis of renal tissue slides [6], [7], [8], [9]. When writing this manuscript, the literature contained no references to computational models specifically focused on the automated classification of podocyte degenerative changes using images of renal glomeruli.

Some authors segmented podocytes for quantitative studies in segmental sclerosis, diabetic nephropathy, and anti-neutrophil antibody-associated glomerulonephritis [10], [11], [12], [13], [14], [15]. Murazesk et al. [13] analyzed podocyte morphological changes and depletion by segmenting them in PAS-stained images. On the other hand, Barros et al. [15] used attribute extraction and convolutional neural networks (CNN) to classify glomeruli images as having or lacking morphological changes in podocytes based on a small dataset with low image diversity.

Although the reports mentioned above aimed to automate the segmentation of podocytes in glomeruli images, this task is not directly useful for making medical decisions about the presence or absence of podocyte degeneration. Instead, classifications are more relevant for this purpose. Moreover, previous studies did not focus on datasets that specifically included podocyte alterations diagnosed by pathologists using light microscopy analysis. Therefore, the models trained on these datasets were not reliably prepared to segment podocytes within glomeruli. Additionally, these datasets lacked diversity in terms of staining techniques. They did not represent real-world images well and did not include other lesions commonly associated with the disease under study. As a result, these datasets provided limited examples compared to those encountered in the routine practice of pathologists.

Motivated by the advances of applying CNN in the field of histology [16], [17] and by the desire to provide the clinical translational aspect, this study aims to fill gaps in computational pathology research on podocytes in renal glomeruli. To this end, we developed an automated binary classification algorithm based on CNN to classify glomeruli images according to the presence of podocyte degeneration. We compared the performance of this classifier to that of three pathologists who evaluated the same images under three

different scenarios. In the first scenario, we assessed the performance of the automated classifier alone. In the second scenario, three pathologists visually analyzed the glomerular images and determined whether podocyte degeneration was present or absent using a standard method. Finally, the pathologists reevaluated the images based on the classifications established by the automated algorithm. We compared the results of each scenario with the level of proteinuria (the nosological diagnosis) related to each image.

As an additional contribution, we created a new dataset of glomeruli images with podocyte degenerative lesions. This dataset consists of 1,143 samples stained with four different techniques (PAS, H&E, PAMS, and trichome). It features a variety of lesions (sclerotic, membranous, and hypercellular) to represent the complexity of images encountered by pathologists in real-world practice. This diverse dataset enhances the reliability of trained models, as they were applied to cases where podocytopathy appeared under different image acquisition conditions and in the context of associated diseases. Finally, this new dataset may also encourage further research on podocytopathy identification.

II. MATERIAL AND METHODS

The experiments were executed in three steps, following a protocol inspired by the work of Ligabue et al. [18], which evaluated the performance of a CNN in classifying immune deposits using immunofluorescence images of renal biopsies. The authors then compared the performance of the CNN against that of a group of three trained pathologists.

For the first step, the goal was to obtain an automated classifier capable of recognizing glomeruli with podocyte degeneration visible by light microscopy. Several automated classifiers (based on CNN architecture) were fitted (trained and validated) using dataset D1. The performance of each model in classifying the images in dataset D2 was ranked to select the most efficient model. In the second step, we assessed the performance of a group of three pathologists who classified the images in dataset D2, to compare their performance against that of the selected computational model. Finally, after a period of 30 days had elapsed (a time interval adopted to avoid classification bias), the same group of pathologists evaluated the images in D2 again, this time with the results from the automated classifier to assess whether this information would influence (i.e., improve) the pathologists' performance. Figure 1 presents an overview of the experiments performed.

A. IMAGE DATA

We assembled two datasets containing digital images of renal glomeruli. Dataset 1 (D1) was used to train and test the automated classifier, while Dataset 2 (D2) was used for performance assessments. Figure 2 illustrates some examples of the images that comprise datasets D1 and D2. The datasets (D1 and D2) include images of glomeruli with various types of podocyte injuries, such as degeneration, hyperplasia, and hypertrophy. This diversity of podocyte injuries provides a

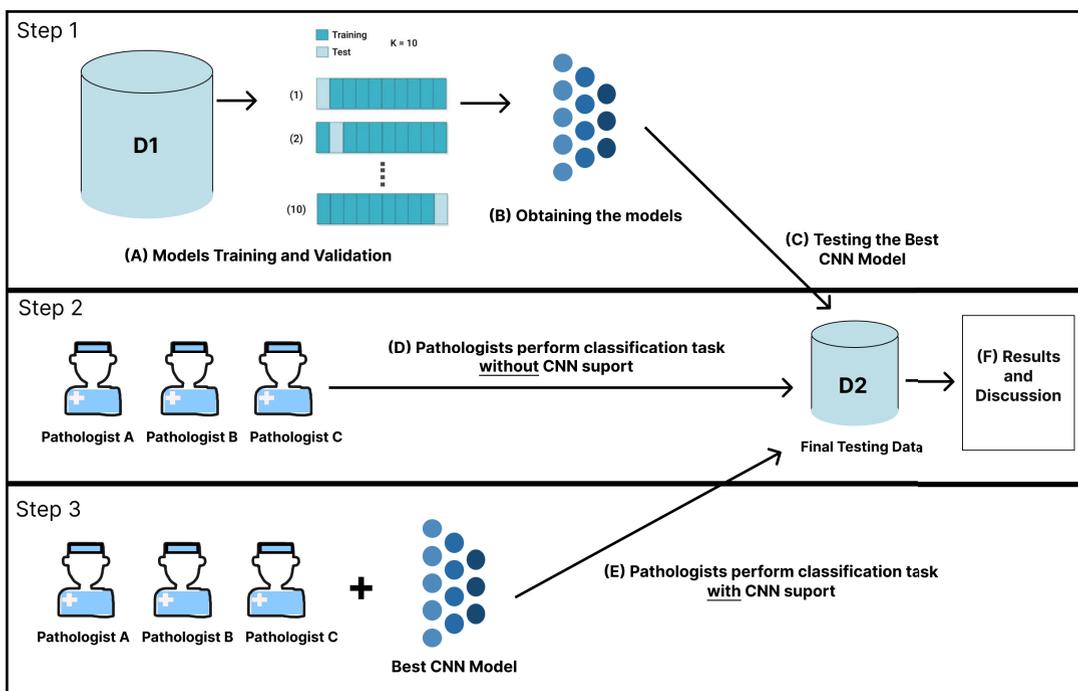


FIGURE 1. Overview of the conducted experiment protocol. In Step 1, several models based on CNN architecture were evaluated to select the one offering the best performance using dataset D2. In Step 2, we assessed three pathologists’ performance in classifying the images in dataset D2. In Step 3, we evaluated the performance of the same pathologists in classifying the images in dataset D2 with the aid of the results obtained from the automated classifier selected in Step 1.

realistic representation of the types of cases encountered in clinical practice, allowing the trained computational models to perform in a manner that reflects the complexity of real-world scenarios. These images did not undergo any pre-processing operations and were delivered to the classifiers exactly as the pathologists provided them. This approach enabled end-to-end processing, as proposed by Goodfellow and colleagues [19].

The images from D1 were obtained by searching image banks provided by two independent pathologists. More than 30,000 images were analyzed, with only 1,043 images selected (513 with and 530 without podocyte degeneration); the selected images employed four different types of staining: PAS, H&E, PAMS, and trichrome (see stain distribution in Table 1). The images originated from the laboratories of different institutions, captured at resolutions ranging from 238×201 to 1920×1440 pixels. In both groups of images (with and without podocyte degeneration), glomeruli contained other unidentified lesions and hypercellular, membranous, and sclerotic lesions labeled by pathologists. The visual inspection of the two independent pathologists who labeled the images in D1 was considered a gold standard used for training the CNN model. Importantly, these two pathologists were not part of the group of three pathologists who performed the classification of dataset D2.

The evaluation dataset (Dataset D2) was created using a different protocol. It consists of 100 images of renal glomeruli (50 with and 50 without podocyte degeneration) stained with PAS, H&E, PAMS, and trichrome. In contrast to Dataset D1,

TABLE 1. Stain distribution in Dataset D1 according to glomeruli with and without podocyte degeneration.

Label	H&E	PAS	Trichome	PAMS	Total
With	44	275	104	90	513 (49%)
Without	64	254	94	118	530 (51%)
Total	108 (10.8%)	529 (50.8%)	198 (18.9%)	208 (19.9%)	1043 (100%)

which was labeled solely based on visual inspection by two pathologists, each image in D2 underwent visual inspection and was annotated with proteinuria data as a gold standard. Furthermore, each image in D2 was obtained from a different patient (100 patients) to ensure robustness in the classification results obtained from computational models. This approach also increased the diversity of cases used to evaluate the CNN model’s performance. The distribution of stains in Dataset D2 is shown in Table 2.

B. AUTOMATED CLASSIFICATION

The final automated classifier was developed by comparing six different models, and the one that performed the best was selected. Python 3.6 [20] was used for software coding, along with Tensor Flow 2.6 [21], Numpy 1.21.0 [22], Scikit-learn [23], and KerasTuner [24] libraries. The experiments were run on a desktop with an RTX 2080TI GPU, 64 GB of RAM, and an AMD Ryzen 3 CPU, all running on Ubuntu 22.04 OS.

Three classical CNN [25] architectures were considered for automated image classification: Inception Resnet-v2 [26],

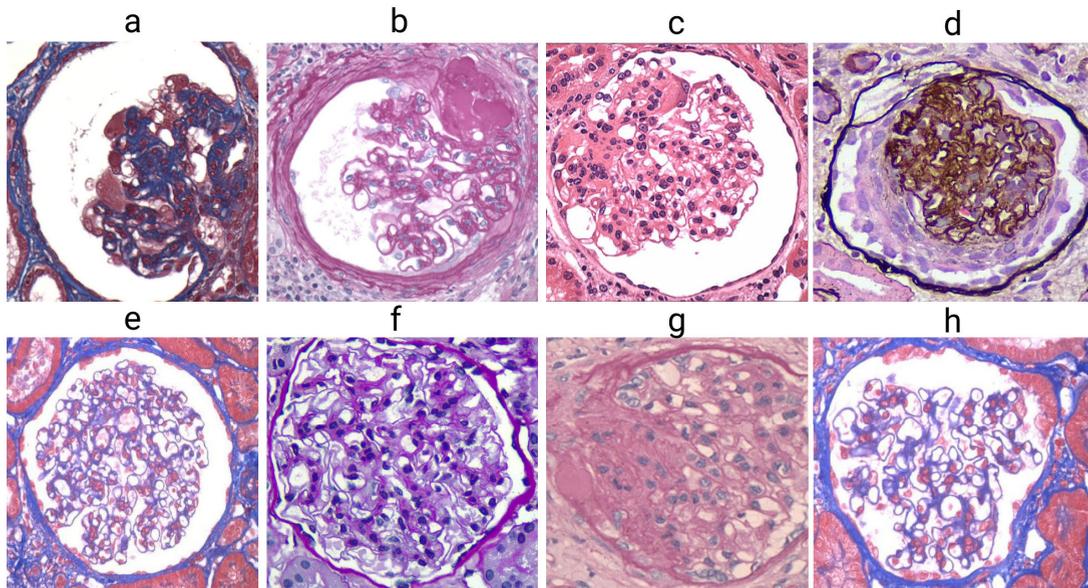


FIGURE 2. Samples from Dataset 1 (D1). Glomeruli with podocyte degeneration (a-d) and glomeruli without podocyte degeneration (e-h).

TABLE 2. Stain distribution in Dataset D2 according to glomeruli with and without podocyte degeneration.

Label	H&E	PAS	Trichome	PAMS	Total
With	18	3	2	17	50 (50%)
Without	12	24	6	8	50 (51%)
Total	30 (30%)	37 (37%)	8 (8%)	25 (25%)	100 (100%)

TABLE 3. Range of values considered during hyperparameter tuning.

Hyperparameters	Assessed values
Batch size	32, 16, 8, 4 and 2
Loss function	Binary crossentropy and Hinge
Optimizer	SGD, Adam, Adagrad and RMSprop
Learning rate	0.1, 0.001, 0.0001 and 0.00001
Number of neurons in dense layers	2048, 1024, 512 and 256

Densenet201 [27], and Efficient-Net B3 [28]. These architectures were chosen due to the provision of reliable results in similar tasks [29], [30], [31] and due to differences in architecture concerning depth (the number of layers is fixed in Densenet201 and Inception Resnet-v2, yet variable in Efficient Net B3) and learning strategies (*e.g.*, residual blocks in Resnet, and dense layers in Densenet201). The three network architectures were trained under two scenarios: A) from scratch (FS) with random weight initialization, and B) using transfer learning (TL) where each network was pre-trained with the ImageNet [32] dataset. This combination resulted in six network models to evaluate.

We fine-tuned the hyperparameters for each CNN model using the hyperband [33], [34] algorithm, which incorporates a Bayesian optimization strategy to expedite the search process through adaptive resource allocation and early stopping techniques prior to training.

To adjust the hyperparameters, in D1, the images were randomly divided into two groups: 70% for training and 30% for testing, a common split percentage for training and validating machine learning models. This split maintained the proportion of images in each class, distinguishing between those with and without podocyte degeneration. This split was exclusively employed for adjusting the hyperparameters of each network. This method, as opposed to a conventional random selection of hyperparameters, yielded an enhanced set,

TABLE 4. Best hyperparameter values of each model.

Model	Bs	Loss	Optimizer	Dense layers	LR
Inception Resnet101 v2 TL	8	Binary Cross-entropy	RMSprop	1024 and 512	0.0001
Densenet201 TL	8	Binary Cross-entropy	Adam	512 and 512	0.001
EfficientNet B3 TL	4	Hinge	RMSprop	Adam and 512	0.0001
Densenet201 FS	4	Binary Cross-entropy	SGD	2024 and 1024	0.001
Inception Resnet 101 v2 FS	8	Binary Cross-entropy	Adam	1024 and 512	0.0001
EfficientNet B3 FS	8	Binary Cross-entropy	Adam	512 and 512	0.0001

contributing to the mitigation of underfitting and overfitting issues.

The hyperband algorithm fine-tuned critical hyperparameters, such as learning rate, optimizer, batch size, loss function, and the number of neurons in the final layers (dense layers) of the network. Table 3 outlines the evaluated values for each hyperparameter. Table 4 shows the results of the best hyperparameter values found for each model candidate.

After obtaining the best combination of hyperparameters offered by the Hyperband algorithm, as depicted in Figure 1, the training of candidate models involved a 10-fold cross-validation. This process included dividing the D1 set into 10 validation subsets. As a result, each of the 6 network combinations underwent training and validation across 10 iterations, utilizing one subset for validation and the remaining subsets for training in each iteration. Consequently, after these iterations, the cross-validation process yielded 10 distinct models. From these models, the one exhibiting the highest F-score was chosen as the optimal model. Subsequently, the chosen model underwent a comparison with the other optimal models in each of the six combinations.

Continuing our efforts to address the challenges of overfitting and underfitting, and with the goal of augmenting sample diversity throughout the model training process, we integrated a data augmentation strategy into the training sets marked in the cross-validation. This was achieved through the implementation of the following operations: 8 rotations at angles ranging from 30 to 310 degrees (at 30-degree intervals), both vertical and horizontal flips, random adjustments of brightness, and contrast variation. These operations were carefully selected to expand the dataset while avoiding any mischaracterization of the morphology of glomeruli images, which could potentially hinder network learning and compromise the overall classifier performance.

As a result of these augmentation techniques, the training sets for each cross-validation round expanded significantly from 1043 to 12516 images, with each original image generating 12 new images. This augmentation process contributes to a more robust and diverse training dataset, fostering improved generalization capabilities of the classifier.

We adopted an early stop strategy to limit training time, in which training was interrupted whenever a sequence of ten epochs did not reduce loss. This is a common practice also designed to prevent overfitting and unnecessary training time. Finally, after cross-validation and obtaining the 6 best models, the D2 data set was used to test each model.

C. PATHOLOGIST CLASSIFICATION

To establish a baseline for CNN model performance comparisons, three pathologists classified the images in D2 as either “with” or “without” podocyte degeneration. This task was performed exclusively by visually inspecting images without any additional information. Importantly, the D2 classification task was carried out by three pathologists who did not participate in preparing or annotating either of the two datasets considered herein (D1 and D2). Each of the three pathologists is also affiliated with a different nephropathology institution.

To mitigate potential classification bias due to immediate re-analysis, a 30-day washout period was implemented before the same three pathologists re-classified the same images using the output from the automated classifier as additional information. In other words, before deciding on a given image, each pathologist was allowed to consult the results obtained from the automated classifier. This approach

TABLE 5. Results of k-fold cross-validation indicated the Inception Resnet101 v2 as the best CNN model.

Model	Acc.(%)	Prec.(%)	Rec.(%)	F1(%)
Inception Resnet101 v2 TL	94.7±2.6	95.5±3.6	92.3±4.3	93.7±2.6
Densenet201 TL	91.3±3.5	92.5±3.8	90.2±7.1	91.1±3.9
EfficientNet B3 TL	91.1±2.4	92.8±3.4	89.0±7.1	90.6±2.8
Densenet201 FS	88.4±2.9	90.6±5.9	86.0±5.8	88.0±3.1
Inception Resnet 101 v2 FS	86.6±3.9	89.3±6.0	82.2±9.7	85.1±4.9
EfficientNet B3 FS	83.3±3.1	84.1±5.2	82.2±7.1	82.8±3.4

allowed us to assess the pathologists’ performance both with and without the support of the computational model, minimizing the risk of over-reliance on the model output.

III. RESULTS AND DISCUSSION

The metrics used to assess the automated classifier and pathologists’ performance were: accuracy, precision, recall, F1-score, and area under the ROC (AUC) curve [35]. In addition to these metrics, we also calculated the reliability of agreement between the three pathologists using the Fleiss’ Kappa [36] metric, precisely to determine whether the inclusion of results from the automated classifier in the image classification task altered inter-observer agreement.

Table 5 lists the average values the CNN models achieve. The model that produced the highest F1-score (96.9%) obtained through 10-fold cross-validation was Inception Resnet101 v2 with transfer learning architecture, offering an overall average of 94.7% of accuracy, 95.5% of precision, 92.3% of recall, and 93.7% of F1-score values. We selected the best one with 96.9% of the F1-score from the ten models obtained by cross-validation.

ROC curves were employed to illustrate the performance of each classifier in conformity with the variability of the discrimination threshold. Ideal/perfect performance is considered when a true positive rate of 1 and a false positive rate of 0 are achieved. Figure 3 shows the ROC curves generated by each model in classifying the images contained in D2. Again, the pre-trained Inception Resnet 101 v2 offered the best discrimination using this dataset, providing an AUC of 0.95 compared to 0.90 from the pre-trained Densenet201. All pre-trained networks surpassed the results obtained by their respective FS versions, which indicates improved performance obtained from using pre-trained networks to classify images of glomeruli with podocyte degenerative changes.

The pathologists’ performance improved when they consulted the results from the automated classifier before providing their classification determinations. Table 6 reveals that the success rate of each pathologist improved independently when using the information obtained from the automated classifier. A comparison of the pathologists’ performance before and after consulting the automated classifier

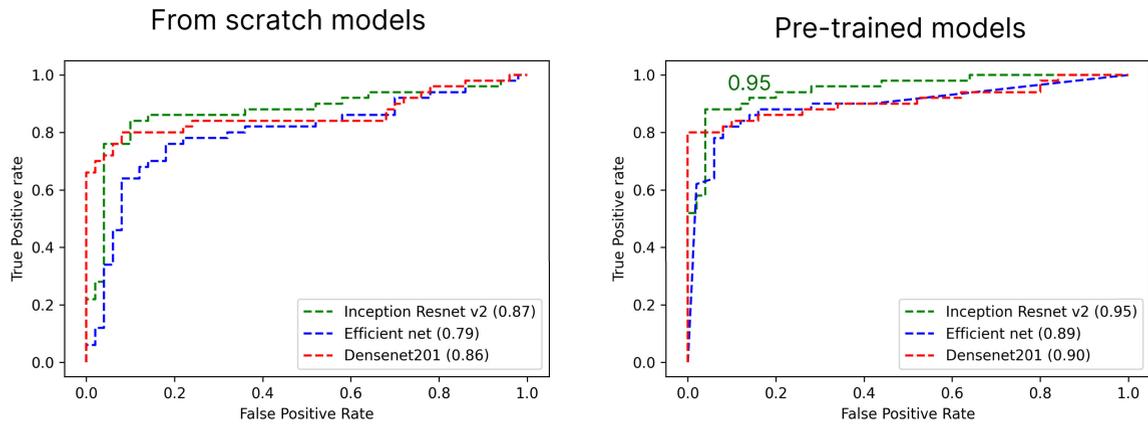


FIGURE 3. Area under the curve (AUC) of the models when classifying images in the D2 dataset. The figure on the left shows a comparison between the models trained with random weight initialization (“from scratch models”) and the same models trained using transfer learning (“pre-trained models”). The pre-trained models presented a greater AUC in all cases, with the Inception ResNet v2 network offering the best results (0.95, green dotted line).

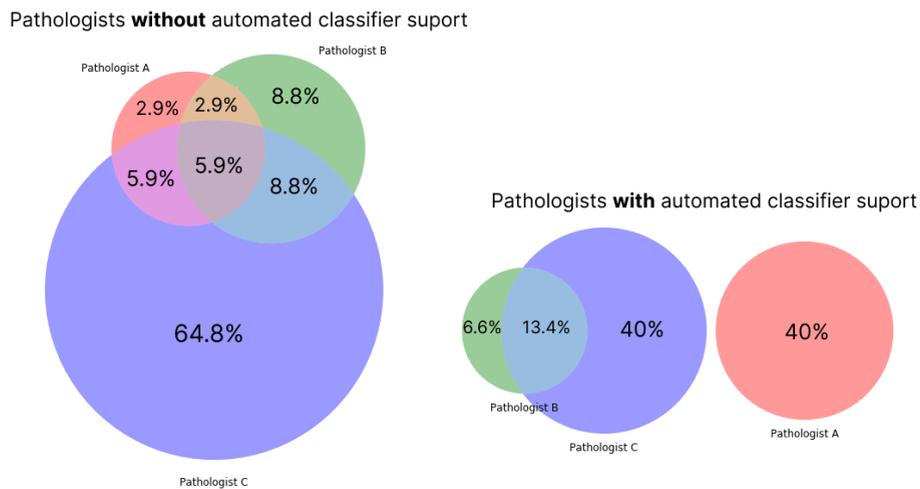


FIGURE 4. Venn diagrams illustrate the distribution of images misdiagnosed by pathologists with and without an automated classifier. Broader distribution and greater disagreement were observed (seven groups in different colors on the left diagram) without the aid of the CNN model.

revealed that overall accuracy increased from $86.3 \pm 10\%$ to $95.0 \pm 2\%$.

Considering the pathologists’ performance in images with and without podocyte degeneration separately, the average accuracy in analyzing images with podocyte degeneration improved from $84.6 \pm 17\%$ to $96.6 \pm 3\%$. In contrast, the average accuracy of images without podocyte degeneration improved from $88.0 \pm 5\%$ to $98.3 \pm 2\%$. Regarding the pathologists’ performance considering images grouped by staining technique: the overall accuracy for PAS-stained images improved from $83.7 \pm 9\%$ to $94.5 \pm 2\%$; for H&E-stained images, from $80.0 \pm 17\%$ to $95.5 \pm 5\%$; yet for trichrome-stained images, no difference was observed as average accuracy remained at $91.6 \pm 7\%$; for PAMS-stained images, a discrete decrease was observed, from $96.0 \pm 6\%$ to $94.6 \pm 4\%$ (within the range of standard deviation).

The automated classifier also contributed to an increase in the degree of agreement (Fleiss’ kappa) between pathologists, which rose from 0.59 (moderate agreement) in their initial

analysis to 0.83 (much better agreement) after introducing the aid of the automated classifier – a significant improvement of approximately 40%. Although Fleiss’ kappa is a metric based on Cohen’s Kappa, it differs in that it assesses agreement between three or more observers. This coefficient was calculated to evaluate whether an automated classifier also influences pathologists’ agreement when visually analyzing images.

The increased concordance observed among the three pathologists is particularly noteworthy, given that they are affiliated with different centers. This finding is significant because research reports [37], [38], [39] have shown that pathologists working in the same center tend to exhibit higher agreement rates. The use of pathologists from different centers allowed us to more rigorously evaluate the efficacy of the automated classifier as a support tool.

Figure 4 presents two Venn diagrams, which illustrate the distribution of errors made by pathologists in their analyses with and without the aid of the automated classifier.

TABLE 6. Performance of three pathologists with and without the aid of an automated classifier in classifying images from Dataset D2. The availability of CNN model results showed an increase in overall classification accuracy.

Classifier	Acc.(%)	Prec.(%)	Rec.(%)	F1(%)	Fleiss' kappa
Pathologist A	94.0	94.0	94.1	94.0	0.59
Pathologist B	91.0	87.0	96.0	91.0	
Pathologist C	74.0	80.0	64.0	90.6	
Pathologist average	88.3 ±10.7	87.0 ±7.0	84.6 ±17.9	91.4 ±12.5	-
CNN	91.0	86.0	95.0	90.9	-
Pathologist A + CNN	96.0	100.0	92.5	96.1	0.83
Pathologist B + CNN	97.0	94.0	100	96.9	
Pathologist C + CNN	92.0	86.0	97.7	91.4	
Pathologist + CNN average	95.0 ±2.6	93.3 ±7.0	96.7 ±3.7	96.1 ±2.9	-

It is evident that, following the introduction of the classifier as a support tool, the sets of images that the pathologists misdiagnosed were different. For example, Pathologists A and C committed errors on entirely different sets of images, indicating a clear change in how the pathologists analyzed the images using the information provided by the classifier.

IV. CONCLUSION AND FUTURE DIRECTION

This study introduces a novel “human-in-the-loop” approach that integrates a convolutional neural network (CNN) with expert pathologist input to accurately identify glomerular podocytopathy. Our findings demonstrate the potential of AI models to augment pathologists’ capabilities in visual analysis of glomeruli with podocyte degeneration. To further contribute to the field, we are making the datasets utilized in this study publicly accessible. By sharing these curated data, we aim to accelerate the development of robust computational models for classifying podocyte degenerative changes in renal glomeruli images.

The limited quantity of training images and the involvement of a relatively small number of pathologists in the annotation process represent significant constraints on the generalizability and robustness of our current models. We acknowledge these limitations and plan to address them in future work by expanding the dataset and involving a larger cohort of pathologists.

We expect that pathologists will be able to adopt automated classifier tools to support the task of image analysis. In our future work, we plan to expand and improve the existing research by incorporating additional datasets to build a multi-class tool capable of performing semantic segmentation of podocytes in glomeruli. This expanded tool will identify and locate podocytes and indicate the specific type of affecting lesion, including normal, degenerative, hyperplastic, or hypertrophic. This comprehensive analysis will provide a more nuanced understanding of podocyte degenerative changes and aid in accurate diagnosis.

Furthermore, we intend to explore and evaluate custom models and other network architectures that may further enhance the performance of the presently developed automated classifier. By investigating alternative deep learning architectures, we aim to improve the accuracy and efficiency of classifying podocyte lesions on conventional slides. This is particularly important as human eyes can only diagnose certain podocyte lesions under electron microscopy, which is time-consuming and resource-intensive. By leveraging advanced network architectures, we anticipate bridging this diagnostic gap and enabling more efficient and accurate diagnosis of podocyte lesions.

Moreover, we recognize the importance of collaboration and data sharing in advancing the field of podocyte degenerative change recognition. Therefore, we plan to collaborate with other research institutions and experts to expand our dataset and ensure a diverse and representative collection of glomerular images. This collaborative effort will facilitate the development of robust and generalizable automated classifier tools that can be applied across different clinical settings and populations.

CODE AND DATA AVAILABILITY

Network training scripts and the best podocyte degenerative changes recognition CNN model will be freely available at GitHub repository. Datasets D1 and D2 are available only to researchers and students for academic purposes. Access can be requested through the Pathospotter research group website (<http://pathospotter.uefs.br/>) or by the following link.

AUTHOR CONTRIBUTIONS STATEMENT

George Oliveira Barros performed all deep neural network experimentation. Angelo Amancio Duarte, Luciano Rebouças de Oliveira, Washington Luis Conrado dos-Santos, and Flavio de Barros Vidal reviewed and validated the experimental protocol and experimental results. Washington Luis Conrado dos-Santos and David Campos Wanderley provided images of renal glomeruli, the ground truth for D1 and D2 datasets, and technical support in the medical field. Henrique Machado de Sousa Proença, José da Silva, and Stanley Almeida Araújo performed the classification of the validation dataset (D2). All authors contributed to the writing and revision of the final manuscript.

DECLARATION of COMPETING 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.

ETHICAL CONSIDERATIONS

The study was conducted in accordance with resolution No. 466/12 of the National Health Council. The patients were not directly involved in the study. The consent was obtained from all subjects and/or their legal guardians, and to preserve confidentiality, the images (including those shown in the paper) were separated from the other patient data. No data

presented herein allows patient identification. All the procedures were approved by the Ethics Committee for Research Involving Human Subjects of the Gonçalo Moniz Institute from the Oswald Cruz Foundation (CPqGM/FIOCRUZ), Protocols No. 188/09 and No. 1.817.574.

ACKNOWLEDGMENT

The authors would like to thank the Gonçalo Muniz Research Institute, Oswald Cruz Foundation (IGM/FIOCRUZ); the State University of Feira de Santana, Instituto Federal Goiano (Campus Posse); and the University of Brasilia for supporting this research with financial assistance, as well as computational and physical infrastructure. They would also like to thank Andris K. Walter for critical analysis, English language revision, and manuscript copyediting assistance.

REFERENCES

- [1] J. B. Kopp et al., "Podocytopathies," *Nature Rev. Disease Primers*, vol. 6, no. 1, p. 68, 2020.
- [2] J. Silvano, "A case of lupus podocytopathy and a review of the literature," *Portuguese J. Nephrol. Hypertension*, vol. 35, pp. 39–42, Apr. 2021.
- [3] A. Diez-Sampedro, O. Lenz, and A. Fornoni, "Podocytopathy in diabetes: A metabolic and endocrine disorder," *Amer. J. Kidney Diseases*, vol. 58, no. 4, pp. 637–646, Oct. 2011.
- [4] F. Ziyadeh and G. Wolf, "Pathogenesis of the podocytopathy and proteinuria in diabetic glomerulopathy," *Current Diabetes Rev.*, vol. 4, no. 1, pp. 39–45, Feb. 2008.
- [5] H. Trimarchi and R. Coppo, "Podocytopathy in the mesangial proliferative immunoglobulin a nephropathy: New insights into the mechanisms of damage and progression," *Nephrol. Dialysis Transplantation*, vol. 34, no. 8, pp. 1280–1285, Aug. 2019.
- [6] G. O. Barros, B. Navarro, A. Duarte, and W. L. C. dos-Santos, "PathoSpotter-K: A computational tool for the automatic identification of glomerular lesions in histological images of kidneys," *Sci. Rep.*, vol. 7, no. 1, pp. 1–8, Apr. 2017.
- [7] W. L. C. dos-Santos, L. A. R. de Freitas, A. A. Duarte, M. F. Angelo, and L. R. Oliveira, "Computational pathology, new horizons and challenges for anatomical pathology," *Surgical Experim. Pathol.*, vol. 5, no. 1, p. 10, Dec. 2022.
- [8] J. Silva et al., "Boundary-aware glomerulus segmentation: Toward one-to-many stain generalization," *Computerized Med. Imag. Graph.*, vol. 100, Sep. 2022, Art. no. 102104.
- [9] P. Chagas et al., "Uncertainty-aware membranous nephropathy classification: A monte-carlo dropout approach to detect how certain is the model," *Comput. Methods Biomech. Biomed. Eng., Imag. Visualizat.*, vol. 11, no. 3, pp. 288–298, May 2023.
- [10] C. Zeng et al., "Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning," *J. Pathol.*, vol. 252, no. 1, pp. 53–64, Jul. 2020.
- [11] D. Govind et al., "PodoSighter: A cloud-based tool for label-free podocyte detection in kidney whole slide images," *J. Amer. Soc. Nephrol.*, vol. 32, no. 11, pp. 2795–2813, 2021.
- [12] D. Govind et al., "Automated detection and quantification of wilms' Tumor 1-positive cells in murine diabetic kidney disease," *Proc. SPIE*, vol. 11603, pp. 76–82, Jan. 2021.
- [13] K. E. Maraszek et al., "The presence and location of podocytes in glomeruli as affected by diabetes mellitus," *Proc. SPIE*, vol. 11320, pp. 298–307, May 2020.
- [14] M. Zimmermann et al., "Deep learning-based molecular morphometrics for kidney biopsies," *JCI Insight*, vol. 6, no. 7, Apr. 2021, doi: [10.1172/jci.insight.144779](https://doi.org/10.1172/jci.insight.144779).
- [15] G. Barros, D. Wanderley, L. Rebouças, W. Santos, A. Duarte, and F. Vidal, "PodNet: Ensemble-based classification of podocytopathy on kidney glomerular images," in *Proc. 17th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2022, pp. 405–412.
- [16] S. Banerji and S. Mitra, "Deep learning in histopathology: A review," *WIREs Data Mining Knowl. Discovery*, vol. 12, no. 1, 2022, Art. no. e1439, doi: [10.1002/widm.1439](https://doi.org/10.1002/widm.1439).
- [17] C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101813.
- [18] G. Ligabue et al., "Evaluation of the classification accuracy of the kidney biopsy direct immunofluorescence through convolutional neural networks," *Clin. J. Amer. Soc. Nephrol.*, vol. 15, no. 10, pp. 1445–1454, Oct. 2020.
- [19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>.
- [20] G. Van Rossum and F. L. Drake Jr., *Python Reference Manual*. Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [21] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Design Implement.*, 2016, pp. 265–283.
- [22] C. R. Harris et al., "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.
- [23] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
- [24] T. O'Malley et al. (2019). *Keras Tuner*. [Online]. Available: <https://github.com/keras-team/keras-tuner>
- [25] J. Salas, F. de Barros Vidal, and F. Martinez-Trinidad, "Deep learning: Current state," *IEEE Latin Amer. Trans.*, vol. 17, no. 12, pp. 1925–1945, Dec. 2019.
- [26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015, *arXiv:1512.00567*.
- [27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [28] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*.
- [29] S. M. Thomas, J. G. Lefevre, G. Baxter, and N. A. Hamilton, "Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101915.
- [30] Y. Li, J. Wu, and Q. Wu, "Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning," *IEEE Access*, vol. 7, pp. 21400–21408, 2019.
- [31] R. Hong, W. Liu, D. DeLair, N. Razavian, and D. Fenyö, "Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models," *Cell Rep. Med.*, vol. 2, no. 9, Sep. 2021, Art. no. 100400.
- [32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 248–255.
- [33] L. Li, K. G. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Efficient hyperparameter optimization and infinitely many armed bandits," 2016, *arXiv:1603.06560*.
- [34] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," *J. Mach. Learn. Res.*, vol. 18, no. 185, pp. 1–52, 2018.
- [35] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*. Cham, Switzerland: Springer, 2013.
- [36] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychol. Bull.*, vol. 76, no. 5, pp. 378–382, Nov. 1971.
- [37] A. M. El-Badry et al., "Assessment of hepatic steatosis by expert pathologists: The end of a gold standard," *Ann. Surg.*, vol. 250, no. 5, pp. 691–697, Nov. 2009.
- [38] A. Angelini et al., "A web-based pilot study of inter-pathologist reproducibility using the ISHLT 2004 working formulation for biopsy diagnosis of cardiac allograft rejection: The European experience," *J. Heart Lung Transplantation, Off. Publication Int. Soc. Heart Transplantation*, vol. 30, pp. 1214–1220, Nov. 2011.
- [39] M. Costantini, "Interobserver agreement in the histologic diagnosis of colorectal polyps the experience of the multicenter adenoma colorectal study (SMAC)," *J. Clin. Epidemiol.*, vol. 56, no. 3, pp. 209–214, Mar. 2003.