Clinical Kidney Journal, 2023, vol. 16, no. 12, 2503–2513

https://doi.org/10.1093/ckj/sfad153
Advance Access Publication Date: 18 July 2023
Original Article

ORIGINAL ARTICLE

Development of a multiple convolutional neural network–facilitated diagnostic screening program for immunofluorescence images of IgA nephropathy and idiopathic membranous nephropathy

Peng Xia[1,*], Zhilong Lv[2,3,*], Yubing Wen[1,*], Baichuan Zhang[4], Xuesong Zhao[1], Boyao Zhang[4], Ying Wang[1], Haoyuan Cui[1], Chuanpeng Wang[1], Hua Zheng[1], Yan Qin[1], Lijun Sun[5], Nan Ye[5], Hong Cheng[5], Li Yao[6], Hua Zhou[7], Junhui Zhen[8], Zhao Hu[9], Weiguo Zhu[10], Fa Zhang[2], Xuemei Li[1], Fei Ren[2] and Limeng Chen[1]

[1]Department of Nephrology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, [3]School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China, [4]Beijing Zhijian Life Technology, Beijing, China, [5]Department of Nephrology, Beijing Anzhen Hospital, Capital Medical University, Beijing, China, [6]Department of Nephrology, First Hospital Affiliated to China Medical University, Shenyang, China, [7]Department of Nephrology, Shengjing Hospital of China Medical University, Shenyang, China, [8]Department of Pathology, Qilu Hospital of Shandong University, Jinan, China, [9]Department of Nephrology, Qilu Hospital of Shandong University, Jinan, China and [10]Department of Information, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

*Peng Xia, Zhilong Lv and Yubing Wen contributed equally to this work.
Correspondence to: Limeng Chen; E-mail: chenlimeng@pumch.cn; Fei Ren; E-mail: renfei@ict.ac.cn

ABSTRACT

Background. Immunoglobulin A nephropathy (IgAN) and idiopathic membranous nephropathy (IMN) are the most common glomerular diseases. Immunofluorescence (IF) tests of renal tissues are crucial for the diagnosis. We developed a multiple convolutional neural network (CNN)-facilitated diagnostic program to assist the IF diagnosis of IgAN and IMN.
Methods. The diagnostic program consisted of four parts: a CNN trained as a glomeruli detection module, an IF intensity comparator, dual-CNN (D-CNN) trained as a deposition appearance and location classifier and a post-processing module. A total of 1573 glomerular IF images from 1009 patients with glomerular diseases were used for the training and validation of the diagnostic program. A total of 1610 images of 426 patients from different hospitals were used as test datasets. The performance of the diagnostic program was compared with nephropathologists.

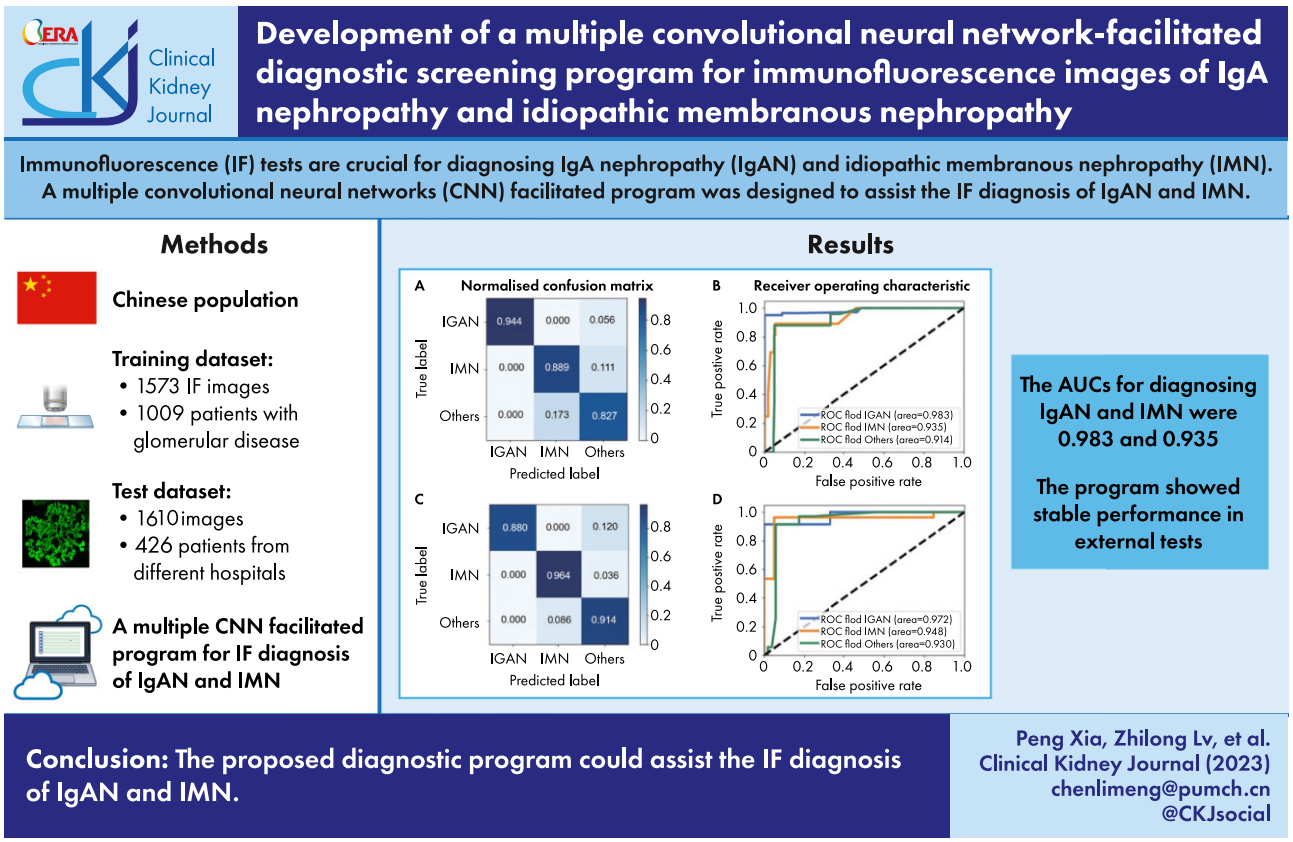
Received: 28.11.2022; Editorial decision: 31.5.2023

Results. In >90% of the tested images, the glomerulus location module achieved an intersection over union >0.8. The accuracy of the D-CNN in recognizing irregular granular mesangial deposition and fine granular deposition along the glomerular basement membrane was 96.1% and 93.3%, respectively. As for the diagnostic program, the accuracy, sensitivity and specificity of diagnosing suspected IgAN were 97.6%, 94.4% and 96.0%, respectively. The accuracy, sensitivity and specificity of diagnosing suspected IMN were 91.7%, 88.9% and 95.8%, respectively. The corresponding areas under the curve (AUCs) were 0.983 and 0.935. When tested with images from the outside hospital, the diagnostic program showed stable performance. The AUCs for diagnosing suspected IgAN and IMN were 0.972 and 0.948, respectively. Compared with inexperienced nephropathologists, the program showed better performance.
Conclusion. The proposed diagnostic program could assist the IF diagnosis of IgAN and IMN.

LAY SUMMARY

A multiple convolutional neural network–facilitated diagnostic program was designed to deliver fast and accurate diagnostic suggestions of immunofluorescence images from immunoglobulin A nephropathy and idiopathic membranous nephropathy (IMN). The accuracy, sensitivity and specificity of diagnosing suspected IgAN were 97.6%, 94.4% and 96.0%, respectively. The accuracy, sensitivity and specificity of diagnosing suspected IMN were 91.7%, 88.9% and 95.8%, respectively. The corresponding areas under the curve were 0.983 and 0.935. When tested with images from the outside hospital, the diagnostic program showed stable performance.

GRAPHICAL ABSTRACT



Keywords: convolutional neural network, idiopathic membranous nephropathy, IgA nephropathy, immunofluorescence

INTRODUCTION

The prevalence of chronic kidney disease (CKD) is increasing worldwide [1]. Glomerulonephritis (GN) is the leading cause of CKD in developing countries [2]. In the Chinese population, immunoglobulin A nephropathy (IgAN) and idiopathic membranous nephropathy (IMN) accounted for >40–60% of primary GN [3–5]. There are no diagnostic serum biomarkers for IgAN. Although the anti-phospholipase A2 receptor (PLA2R) antibody is specific for diagnosing IMN, ≈30% of IMN cases test negative [6]. Microscopic haematuria and non-nephrotic-range proteinuria

are also seen in IMN [7, 8]. Therefore, renal biopsy remains an irreplaceable method to diagnose IgAN and IMN.

Immunofluorescence (IF) tests are necessary for pathologic diagnoses of IgAN and IMN, which indicates locations and appearances of the deposits, such as immunoglobulins, light chains and complements in renal tissues [9, 10]. In China, experienced nephropathologists are lacking. A tool of enough efficiency and accuracy is needed to assist in the diagnosis of common GNs such as IgAN and IMN.

Deep learning methods represented by neural network models are powerful representational learning models and have been applied to fields of image analysis [11]. The convolutional neural network (CNN) is one of the examples applied in image classification and segmentation, where stacked multiple convolution layers can amplify aspects of the input and suppress irrelevant variations [12]. The application of CNN in image analysis of kidney histopathology has proved to be feasible for its excellent feature learning ability in computer vision tasks [13–22]. Most of the CNN-based studies of kidney disease have mainly focused on structural segmentation [15, 17–19]. Some recent studies have used CNN for classifications of morphology, location and the appearance of deposits in IF images from patients with renal diseases [23–25]. However, attempts to apply a CNN-based diagnostic program to facilitate IF diagnosis of glomerular diseases remain lacking.

In this study we established a multiple CNN-facilitated diagnostic program based on the training and validation of 1573 glomerular IF images from 1009 patients and tested the program prospectively in 426 patients from different hospitals to evaluate its diagnostic efficacy versus nephropathologists.

MATERIALS AND METHODS

IF images dataset used for the development of the diagnostic program

A total of 1573 glomerular IF images from 1009 patients with glomerular diseases diagnosed between January 2016 and December 2020 were collected from Peking Union Medical College Hospital (PUMCH). These images were used for training and validation of the glomeruli detection module, the IF deposition location and appearance classification, as well as the diagnostic program (Fig. 1a and Table 1). This study was approved by the Institutional Review Boards of PUMCH (S-K913 and S-K1421) and other participating centers. The informed consent was waived since de-identified images were used.

IF images dataset used to test the diagnostic program

The IF images from patients who underwent pathological investigations at PUMCH from March 2021 to December 2021 were used to test the program. After excluding the blurred images, 1171 images from 338 patients were used as test datasets (Table 1 and Supplementary data, Fig. S1). Another 439 IF images from 88 patients were provided by Beijing Anzhen Hospital for external tests. These images were original pictures taken from the microscope without pre-processing.

Data annotation

Marks leading to personal information identification or interfering with CNN training were erased from the IF images. For the training of the glomeruli location module, the images were annotated for the location of the glomerulus with a rectangle using labelImage (Supplementary data, Fig. S2). The ground truth for intensity, deposition location and appearance, as well as the suggested diagnosis of all the IF images, was based on the consensus of diagnosis by two nephropathologists (Yubing Wen and Lin Duan) from PUMCH, who both had >20 years of work experience.

Diagnostic program design

This program included four modules: a glomeruli detection module, an IF intensity comparator, a deposition location and appearance classifier and a post-processing module.

Glomeruli detection module

Training and validation datasets

A CNN was trained using 253 IF images containing one or more glomerulus (Supplementary data, Fig. S1) to detect the glomeruli (Fig. 1b). The training dataset consisted of 201 images and the remaining 52 images were used for validation.

Development process

We trained a CNN based on the YOLOv5 frame, a single-stage target location algorithm (Fig. 2). The trained CNN first initialized several bounding boxes. The bounding box with the highest intersection over union (IOU) overlapping with the ground truth label was selected as the output (Supplementary data, Fig. S3).

IF intensity comparator

Dataset for development

A total of 388 IF images with an intensity ≥2+ were used to develop the intensity comparator (Fig. 1c).

Development process

The IF intensity of a certain image was calculated by the mean grey value:

$$\text{Mean Grey Value} = \frac{\sum_{i=1}^{Pixel} \text{Grey Value}}{\text{Pixel}}$$

The Otsu threshold method was used to extract the feature of grey value in IF images. The images were divided into background and foreground. The target was to maximize the variance between the grey value of background and foreground images. Morphological opening-and-closing operations were used to repair the edge region to extract the grey values of the target region. Finally, the mean grey values of each image were calculated and sorted. The cut-off gray value was set at 55 since 95% of the IF images with intensity ≥2+ showed greater gray value.

Classification of IF deposition appearance and location

Training and validation datasets

A total of 1320 images consisting of IgAN, IMN and other kinds of GNs were selected for the training dataset (1216 images) and validation dataset (104 images) (Fig. 1d). These images were classified into three categories: irregular granular mesangial deposition, which were all images of IgAN; fine granular deposition along the glomerular basement membrane (GBM),

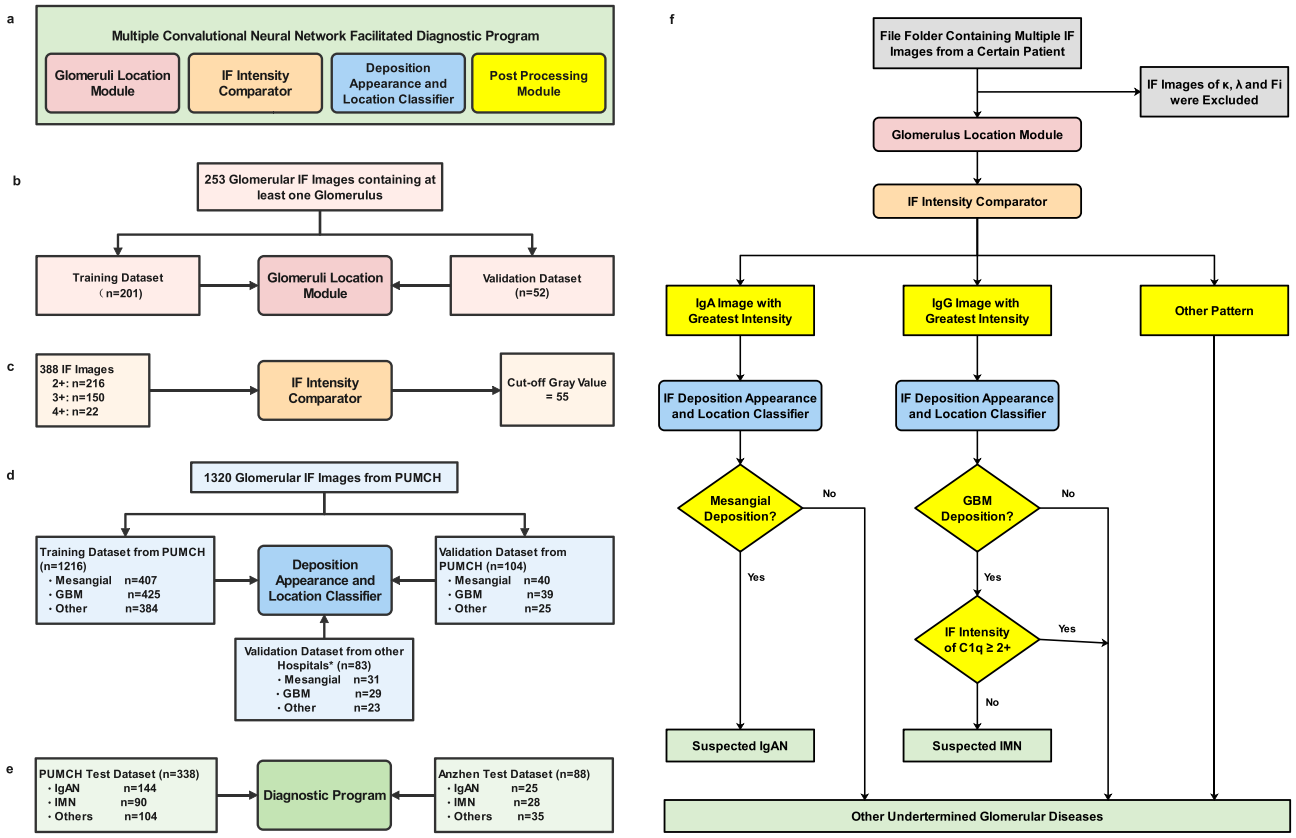


Figure 1: Study design. **(a)** The diagnostic program consisted of four parts: a CNN trained as a glomeruli location module that could localize IF-positive glomeruli, an IF intensity comparator that could identify the image with the greatest intensity, a D-CNN trained as a deposition appearance and location classifier and an additional decision-adjusting module that could adjust the final output. **(b)** A total of 253 IF images containing at least one glomerulus were used for training and validation of the glomerulus location module. **(c)** A total of 388 IF images of different intensities were used to generate the cut-off value of 2+ intensity. **(d)** A total of 1320 images consisting of 447 IgAN, 464 IMN and 409 images of other GNs were used as the training dataset (1216 images) and validation dataset (104 images) for the deposition appearance and location classifier. Another 83 IF images from other hospitals were also used for validation. **(e)** IF images from 338 patients diagnosed at PUMCH were used to test the performance of the diagnostic program. Another dataset of 88 patients from Anzhen Hospital was also used for the test. **(f)** The suggested diagnosis was divided into three categories: suspected IgAN, suspected IMN and undetermined glomerular diseases. All the positive IF images of different immunoglobins and complement components from a certain patient are input into the diagnostic program. After the glomerulus location and IF intensity comparison, the IgG or IgA image with the greatest intensity is input into the deposition appearance and location classifier (D-CNN). Otherwise, the diagnostic program will output undetermined glomerular diseases, which requires further investigation by an experienced nephropathologist. If the IgA image showed typical mesangial deposition, the diagnosis of suspected IgAN was suggested. If the IgG image showed typical GBM deposition and the intensity of C1q deposition was ≤2+, the diagnosis of suspected IMN was suggested. All other cases were categorized as undermined glomerular diseases. *Other hospitals: First Hospitals Affiliated to Chinese Medical University, Shengjing Hospital of Chinese Medical University, Beijing Anzhen Hospital, Capital Medical University and Qilu Hospital, Shandong University.

which were all images of IMN; and other deposition patterns (Supplementary data, Fig. S4). An additional 83 IF images from different hospitals (the First Hospitals Affiliated to Chinese Medical University, Shengjing Hospital of Chinese Medical University, Beijing Anzhen Hospital, Capital Medical University and Qilu Hospital, Shandong University) were provided for external validation.

Development process

We proposed a dual CNN (D-CNN) model based on the fusion of global and local features to identify the deposition location and appearance classification (Fig. 3). It consisted of a two-branch features extraction module based on pre-trained Inception v3 modules for features extraction and a fully connected layer for features fusion [26]. In the two-branch features extraction module, the upper branch Inception v3 network can extract the global coarse-grained features from the IF images. Meanwhile, the other Inception v3 network is used to extract fine-grained features from three regions, with the high-intensity value determined by the distance transform IF images, which can approximately describe the density of deposits. After multiple convolutions, pooling and other operations of the upper Inception v3 model, the IF images were transformed into 1×32 global feature maps. The input of the lower branch Inception v3 network includes three local regions of the IF images with a size of 150×150 according to the density of deposits approximated by the distance transform images (Supplementary data, Fig. S5). At the end of the D-CNN model, the coarse-grained features and fine-grained features were combined by fully connected layers to predict the classification labels. The data augmentation on the training images includes flip operation, rotation operation and affine transformation (Supplementary data, Fig. S6).

Post-processing module

The program generated three suspected diagnosis categories: suspected IgAN, suspected IMN and undetermined glomerular

Table 1: The number of IF images from PUMCH used for the training, validation and testing

	Images used as training and validation datasets		images used as test datasets	
Diagnosis	Patients (*n*)	IF images (*n*)	Patients (*n*)	IF images (*n*)
IgAN	342	508	122	407
Idiopathic membranous GN	323	544	90	276
LN	260	408	53	268
Post-infectious GN	30	36	1	5
Atypical membranous GN	26	38	29	107
MPGN	23	34	1	3
Henoch–Schönlein purpura	2	2	22	58
Amyloidosis	2	2	3	11
Cryoglobulinaemic GN	1	1	–	–
DN	–	–	13	22
Anti-GBM disease	–	–	3	8
IMN complicated with IgAN	–	–	1	6
Total	1009	1573	338	1171

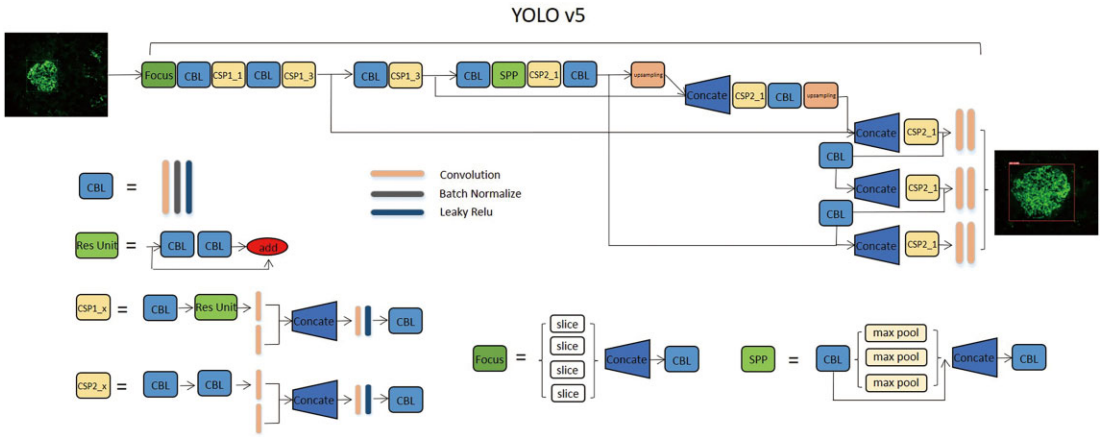


Figure 2: Overview of the CNN trained as a glomerulus location module. We trained a CNN based on the YOLO v5 frame, which was a single-stage target location algorithm. Our purposed CNN first initialized several bounding boxes, which were compared with the ground truth label. IOU was calculated as the percentage of each bounding box overlapping with the ground truth label. The bounding box with the maximum IOU is selected as the output.

diseases. Positive IF images of different immunoglobins and complement components were input into the diagnostic program. After the glomeruli detection and IF intensity comparison, only the IgG or IgA images with the greatest intensity were input into the deposition appearance and location classifier to improve the diagnostic efficiency. Otherwise, the diagnostic program directly output undetermined glomerular diseases, requiring investigation by an experienced nephropathologist (Fig. 1f).

The criteria for suspected IgAN were as follows: the intensity of the IgA-positive image was the greatest among all the positive images and the IgA-positive image was classified as mesangial deposition.

The criteria for suspected IMN were as follows: the intensity of the IgG-positive image was the greatest among all the positive images, the IgG-positive image was classified as GBM deposition and C1q-negative or intensity <2+ since positive C1q deposition along the GBM suggested secondary MN such as lupus nephritis (LN) [27, 28].

The tested IF images were categorized as undetermined glomerular diseases if matching any of the following criteria: negative IF test result of both IgA and IgG, the IF image with the intensity was neither IgA nor IgG, the IgG or IgA image was classified as other deposition patterns by the deposition appearance and location classifier and the IF intensity of C1q was ≥2+.

Program setup and operation process

An online program incorporating the above four modules was developed. All the positive IF images of a certain patient were put into one folder named after the pathology identification. Each image was named as the deposition category, such as IgG, IgA, C3, C1q etc. The whole folder was then input into the program to generate the suggested diagnosis (Supplementary data, Item S1).

Assessments of performance of the proposed diagnostic program

Assessments of performance of the glomeruli detection module

The average IOU was calculated and an IOU ≥0.8 was considered accurate detection.

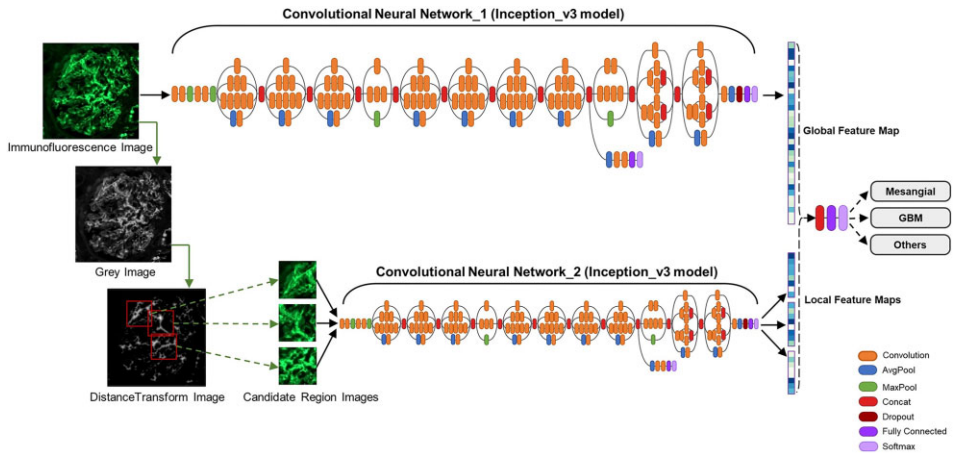


Figure 3: Overview of the D-CNN trained as an IF deposition location and appearance classifier. The D-CNN model consists of two Inception v3 modules for feature extraction and a fully connected layer for feature fusion. The upper Inception v3 model can extract the 1 × 32 global feature maps from the complete original IF images. The lower Inception v3 model is responsible for extracting three 1 × 8 local feature maps from three local images cropped from the original IF images. At the end of the D-CNN model, the fully connected layer is used for the fusion of the 1 × 32 global feature maps and three 1 × 8 local feature maps and finally predicts the best classification labels for the input images.

Assessments of the performance of classification of IF deposition location and appearance

The performance of the deposition appearance and location module was evaluated using accuracy, sensitivity, specificity, F1 score and fallout.

Accuracy is the ratio between the correct predictions and the total predictions:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positive + False\ Negatives}$$

Sensitivity has the same statistical meaning as recall:

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Precision has the same statistical meaning as positive predictive value:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

F1 score is calculated as:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Fallout has the same statistical meaning as the false positive rate:

$$Fallout = \frac{False\ Positives}{True\ Positives + False\ Positives}$$

The receiver operating characteristic (ROC) curve was used to evaluate the accuracy of D-CNN in deposition appearance and location classification. Its performance was compared with traditional state-of-the-art CNNs, which were general CNN models with approximate model complexity [26, 29–31].

Assessment of the performance of the diagnostic program

The performance of the diagnostic program was evaluated for accuracy, sensitivity (recall), specificity and false positive rate (fallout). The calculation of accuracy, sensitivity and false positive rate was the same as listed above.

Specificity is calculated as:

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives}$$

The test results of the diagnostic program in diagnosing IgAN and IMN were compared using Fisher's precise test between data from PUMCH and Beijing Anzhen Hospital.

Performance comparisons between the diagnostic program and nephropathologists

The same test dataset from PUMCH for the diagnostic program was used to test two junior nephropathologists (Haoyuan Cui and Ying Wang) independently. H.C. had 4 years of work experience in IF image interpretation and Y.W. had only 6 months of related experience. Their accuracy, sensitivity, specificity and F1 score were calculated. Cohen's kappa statistic was used for measuring interobserver agreement between pathologists. Their time cost was also documented.

RESULTS

Performance of the glomeruli location module

For the validation dataset of glomeruli location, the mean IOU was 0.83 ± 0.44. The IOUs were >80% in 47 images (90.4%) (Supplementary data, Fig. S7).

Table 2: D-CNN outperformed traditional CNN

Method	Overall accuracy (%)
VGG	78.85
Inception v3	83.65
ResNet50	80.77
DenseNet	81.73
D-CNN	89.42

VGG: Visual Geometry Group; ResNet: Residual Network; DenseNet: Dense Network.

The cut-off value of the IF intensity comparator

The grey value distribution of IF images with intensity ≥2+ are shown in Supplementary data, Fig. S8. The grey value ranged from 11 to 235. The cut-off grey value for 2+ intensity was set at 55 since 95% of these IF images showed a greater grey value.

Performance of the deposition appearance of location classifier (D-CNN)

The D-CNN outperformed traditional CNN in classification of IF deposition appearance and location (Table 2). For the validation dataset from PUMCH, the accuracy, sensitivity, specificity, precision and F1 score of the D-CNN model in recognizing mesangial deposition and GBM deposition were 96.1% and 93.2%, 92.5% and 89.7%, 98.4% and 95.3%, 97.4% and 92.1%, and 0.948 and 0.909, respectively (Table 3 and Fig. 4a). The corresponding AUCs were 0.994 and 0.904 (Fig. 4b). For external tests, the accuracy, sensitivity, specificity, precision and F1 score of recognizing mesangial deposition and GBM deposition were 97.5% and 92.8%, 93.5% and 89.6%, 100% and 94.4%, 100% and 89.7%, and 0.967 and 0.896, respectively (Table 3 and Fig. 4c). The corresponding AUCs were 0.995 and 0.923 (Fig. 4d).

Performance of the diagnostic program

For the test dataset from PUMCH, the accuracy, sensitivity and specificity of diagnosing suspected IgAN were 97.6%, 94.4% and 96.0%, respectively (Table 4). The accuracy, sensitivity and specificity of diagnosing suspected IMN were 91.7%, 88.9% and 95.8%, respectively (Table 4). The confusion matrix is presented in Fig. 5a. The corresponding AUCs were 0.983 and 0.935 (Fig. 5b). For external tests, the accuracy, sensitivity and specificity of diagnosing suspected IgAN were 96.6%, 88.0% and 95.4%, respectively. The accuracy, sensitivity and specificity of diagnosing suspected IMN were 95.4%, 96.4% and 98.2%. The confusion matrix is presented in Fig. 5c. The corresponding AUCs were 0.972 and 0.948 (Fig. 5d). The program demonstrated no significant differences in diagnosing suspected IgAN ($P = .210$) and suspected IMN ($P = .456$) between the data from PUMCH and Beijing Anzhen Hospital.

Performance of the diagnostic program was better than that of inexperienced nephropathologists

The PUMCH test dataset was used to test two junior nephropathologists. Their accuracy, sensitivity, specificity and F1 score for diagnosing suspected IgAN were 88.8% and 97.0%, 73.8% and 93.8%, 83.6% and 97.5%, and 0.848 and 0.993, respectively. The accuracy, sensitivity, specificity and F1 score for diagnosing suspected IMN were 85.8% and 87.3%, 61.1% and 77.8%, 87.0%, and 91.8%, and 0.696 and 0.765, respectively. The Cohen's kappa for suspected IgAN and IMN wase 0.801 and 0.782, respectively (Table 5). It only took 14 minutes for the diagnostic program to finish the test, which accounted for only 14.2% of the time cost (mean time 98.5 minutes) of the junior nephropathologists.

DISCUSSION

GN is still one of the leading causes of CKD in developing countries [1, 32]. Recent studies have shown that IgAN and IMN account for ≈80% of the IF-positive primary GN [4, 5]. The diagnosis confirmation of IgAN and IMN relied on kidney pathological studies and IF assessments were crucial [9, 10]. In this study, we developed a multiple CNN-facilitated diagnostic program to make an IF diagnosis of suspected IgAN and IMN, thus providing rapid triage of glomerular IF images.

The first task of the proposed diagnostic program was to locate the glomeruli. Zhang *et al.* [24] used UNet++ to segment glomeruli from the background in the IF image. Our proposed CNN can simultaneously locate multiple target glomeruli in a single image, preserving the intact glomeruli for IF intensity evaluation.

The second task was to determine the IF deposition with the greatest intensity. Ligabue *et al.* [23] used CNN to identify the IF intensity of different glomerular images but failed to show good agreement with the ground truth. Since the cameras and light sources of fluorescence microscopies in different

Table 3: The D-CNN showed good performance in deposition appearance and location classification

	Validation dataset from PUMCH ($n = 104$)			Validation dataset from other hospitals[a] ($n = 83$)		
Test	Mesangial ($n = 40$)	GBM ($n = 39$)	Other ($n = 25$)	Mesangial[b] ($n = 31$)	GBM[b] ($n = 29$)	Other[c] ($n = 23$)
Accuracy (%)	96.1	93.2	89.4	97.5	92.8	90.3
Sensitivity (%)	92.5	89.7	84.0	93.5	89.6	87.0
Specificity (%)	98.4	95.3	91.1	100	94.4	91.7
Precision (%)	97.4	92.1	75.0	100	89.7	80.0
Fallout (%)	2.6	7.9	25.0	0	10.3	20.0
F1 score	0.948	0.909	0.792	0.967	0.896	0.833

Mesangial: irregular granular mesangial deposition; GBM: fine granular deposition along the GBM.

[a]Other hospitals: First Hospital Affiliated to Chinese Medical University, Shengjing Hospital of Chinese Medical University, Beijing Anzhen Hospital, Capital Medical University, and Qilu Hospital, Shandong University.

[b]All the images of mesangial deposition and GBM deposition used for validation were from Beijing Anzhen Hospital, Capital Medical University.

[c]The images of 'other deposition pattern' used for validation were from First Hospital Affiliated to Chinese Medical University, Shengjing Hospital of Chinese Medical University, Beijing Anzhen Hospital, Capital Medical University, and Qilu Hospital, Shandong University.

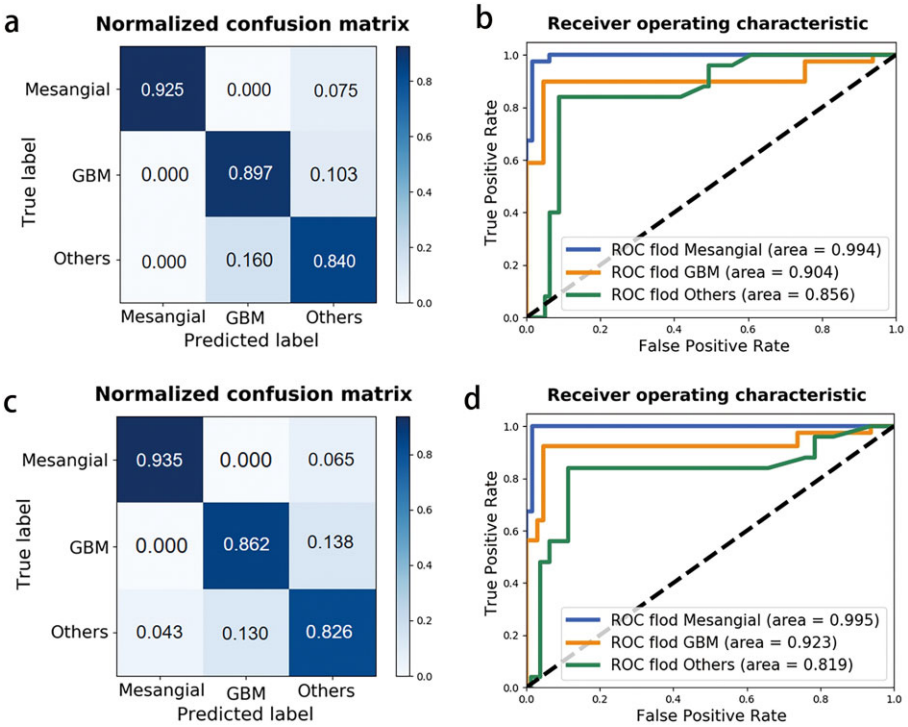


Figure 4: The D-CNN showed good performance in IF deposition location and appearance classification. **(a)** The confusion matrix of the D-CNN when tested with the dataset from PUMCH. The sensitivity of recognizing mesangial deposition, GBM deposition and other deposition patterns was 92.5%, 89.7% and 84.0%, respectively. **(b)** ROC curve of D-CNN when tested with datasets from PUMCH. The AUC for the corresponding three deposition classifications was 0.994, 0.904 and 0.856, respectively. **(c)** The confusion matrix of the D-CNN when tested with datasets from other hospitals. The sensitivity of recognizing mesangial deposition, GBM deposition and other deposition patterns was 93.5%, 89.6% and 87.0%, respectively. **(d)** The ROC curve of D-CNN when tested with datasets from other hospitals. The AUC for the corresponding three deposition classifications was 0.995, 0.923 and 0.819, respectively. Other hospitals: First Hospitals Affiliated to Chinese Medical University, Shengjing Hospital of Chinese Medical University, Beijing Anzhen Hospital, Capital Medical University and Qilu Hospital, Shandong University.

Table 4: The proposed diagnostic program showed good performance in diagnosing IF images from IgAN and IMN

	Test dataset from PUMCH			Test dataset from Beijing Anzhen Hospital		
Test	IgAN ($n = 144$)	IMN ($n = 90$)	Undetermined glomerular diseases ($n = 104$)	IgAN ($n = 25$)	IMN ($n = 28$)	Undetermined glomerular diseases ($n = 35$)
Accuracy (%)	97.6	91.7	89.3	96.6	95.4	92.0
Sensitivity (%)	94.4	88.9	82.7	88.0	96.4	91.4
Specificity (%)	96.0	95.8	92.3	95.4	98.3	94.2
Precision (%)	100	81.6	82.7	100	90.0	88.9
Fallout (%)	0	18.4	17.3	0	10.0	11.1
F1 score	0.971	0.851	0.827	0.936	0.931	0.901

Table 5: The performance of the diagnostic program was better than junior nephropathologists

	Junior nephropathologist 1				Junior nephropathologist 2				
Diagnosis	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 score	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 score	Cohen's Kappa
Suspected IgAN ($n = 144$)	97.0	93.8	97.5	0.993	88.8	73.6	83.6	0.848	0.809
Suspected IMN ($n = 90$)	87.3	77.8	91.8	0.765	85.8	61.1	87.0	0.696	0.782
Undetermined glomerular diseases ($n = 104$)	86.7	80.8	91.7	0.789	74.6	87.5	92.5	0.679	0.671

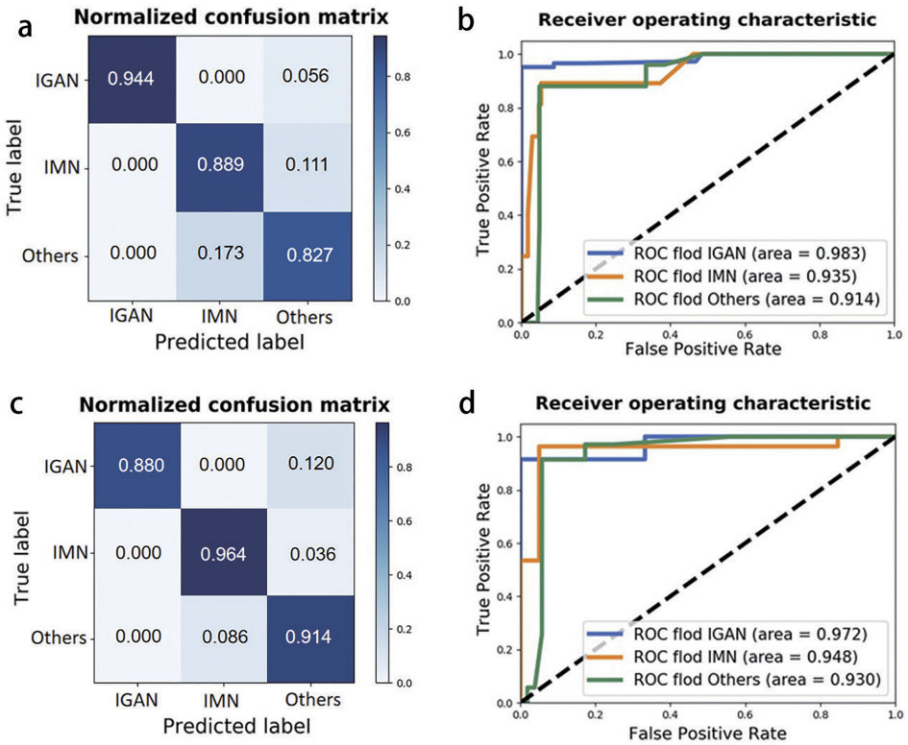


Figure 5: The performance of the diagnostic program in identification of suspected IgAN and IMN. **(a)** The confusion matrix of the diagnostic program when tested with datasets from PUMCH. The sensitivity of diagnosing IgAN, IMN and undetermined glomerular diseases was 88.0%, 96.4% and 91.4%, respectively. **(b)** The ROC curve of the diagnostic program when tested with datasets from PUMCH. The AUC for diagnosing IgAN, IMN and undetermined glomerular diseases was 0.983, 0.935 and 0.914, respectively. **(c)** The confusion matrix of the diagnostic program when tested with a dataset from Beijing Anzhen Hospital. The sensitivity of diagnosing IgAN, IMN and undetermined glomerular diseases was 94.4%, 88.9% and 82.7%, respectively. **(d)** The ROC curve of the diagnostic program when tested with datasets from Beijing Anzhen Hospital. The AUC for diagnosing IgAN, IMN and undetermined glomerular diseases was 0.972, 0.948 and 0. 930, respectively.

institutions are not the same, a trained CNN for the intensity evaluation might not deliver satisfactory generalizability. Therefore we adopted a simple strategy to calculate the grey value of IF images.

The third task was to classify the IF images into different appearances and deposition patterns. We noticed that experienced nephropathologists can accurately recognize different classifications based on the observation of the overall IF images and the minor local regions. It was like fine-grained image categorization, such as bird species classification according to the beaks and feet [33]. We proposed a D-CNN model based on global and local feature fusion. The test results indicated that it could recognize irregular granular mesangial-dominant deposition and fine granular deposition along the GBM very well.

Our final post-processing module was designed to combine the information on deposition patterns with deposition categories to generate the suggested diagnosis.

The proposed diagnostic program showed better diagnostic performance for IgAN than IMN and undetermined glomerular diseases. This was because IgAN had a quite distinct IF deposition pattern compared with the other two groups. For IMN patients, PLA2R antigen staining in biopsy samples is of diagnostic value for PLA2R-related MN [33]. However, PLA2R antigen might also be detected in hepatitis B virus–associated MN in Chinese patients [34]. The IgG subtype might be useful in differentiating SMN from IMN since IgG4 is the prevailing subclass associated with IMN [33]. We believe that the combination of test results of PLA2R antigen staining and IgG subtype might have the potential to improve the diagnostic accuracy of IMN for difficult IF images.

The core of this program is to identify typical IF deposition patterns of IgAN and IMN. Although we tried to train the program to differentiate IF images with only GBM deposition from IF images with GBM and mesangial deposition, 11.1% (10/90) of the tested IMN patients were still misdiagnosed as undetermined glomerular diseases. More training data containing other types of glomerular diseases such as diabetic nephropathy (DN), membranoproliferative GN (MPGN) and anti-GBM disease should be able to improve the CNN performance in the future.

Artificial intelligence has been explored for its potential applications in manipulating data of huge sizes and various forms in the medical area [35]. Different groups have used deep learning networks to locate glomeruli in histopathological images [20, 21]. Kannan *et al.* [15] and Kolachalama *et al.* [19] used CNN to discriminate glomerular sclerosed glomeruli from normal or partially sclerosed glomeruli and tested the relevance between CNN classifications of kidney fibrosis and different degrees of proteinuria, eGFR and CKD stages. Another study showed that CNNs are capable of glomerular segmentation and classification of DN [18].

For IF images, early attempts were to detect glomeruli on IF images of renal tissues [36]. An Italian group trained CNN to identify the appearance, distribution, location and intensity of glomerular IF images [23]. However, their reported CNN was not able to differentiate IF images with both mesangial and GBM deposition. Zhang *et al.* [24] used multiple CNNs to identify

the IF deposition appearance and location based on segmented glomerular IF images. In their study, the deposition appearance and location seemed to be considered as two independent characteristics. Pan *et al*. [25] attempted to use CNN to identify IF images from several glomerular diseases, including IgAN, IMN, LN and DN. For DN and LN, the IF deposition could be highly variable and should be interpreted in combination with light microscopic studies [37]. Sometimes, linear deposition along the GBM may be seen in DN, which could be quite similar to anti-GBM disease [37]. Compared with these reported CNNs, our final proposed diagnostic program demonstrated several potential advantages: it could process IF images containing multiple glomeruli, the IF images photographed could be used without any pre-processing and human knowledge of renal pathology was incorporated into the development of the CNN.

A major problem of applying a CNN in clinical use is the lack of interpretability. It is almost impossible to precisely understand how a network approximates a particular function [14]. A very small variation in the image might affect the output of the CNN, making it a poor generalizer [38]. The most complicated module in our diagnostic program is the D-CNN. To enhance the robustness of this model, a candidate regions localization module was used to extract both global features and local fine-grained features, which could avoid focusing too much on non-discriminative or low-discriminative regions. When tested with images from different hospitals, our diagnostic program demonstrated comparable performance (Tables 3 and 4 and Figs. 4 and 5), suggesting its great potential to generalize.

So far, the reported CNNs in the field of kidney pathology have demonstrated comparable performance, if not superior, to that of nephropathologists [13, 17–19, 23]. Our diagnostic program outperformed two inexperienced nephropathologists in accuracy and time. Given the scarcity of experienced nephropathologists working in remote Chinese hospitals, the test results indicated that this program could enhance the triage of IF images, liberating nephropathologists from repetitive work requiring great attention [13].

Our study has several limitations. The sample size was limited. The program could not identify images with blurry parts, dye contamination and deposits fusion very well. Only part of the included IMN patients had PLA2R antigen staining and IgG subtype data available. At this stage, the program is not fully ready for clinical use because of the limited sample size and the focused disease categories (IgAN and MN). It is necessary to expand the diagnosing ability to more complex IF deposition patterns and constellations, such as MPGN, LN with class V and other types of GN. And it is important to say that IF images alone are not enough for a complete diagnosis. The development of a CNN-assisted program combining the data of IF images, light microscopic images and electron microscopic images should be the future direction.

In conclusion, we developed an automated diagnostic program for IF diagnosis screening of IgAN and IMN that showed the potential for triage of glomerular IF images from GN patients. It consisted of different modules, including the glomeruli location module, IF intensity comparator, IF deposition appearance and location classifier and post-processing module. This program showed stable performance when tested by datasets from different hospitals. It also demonstrated better performance than junior nephropathologists.

SUPPLEMENTARY DATA

Supplementary data are available at *ckj* online.

ACKNOWLEDGEMENTS

We wish to thank Haochen Jiang for his support in designing the online test for nephropathologists. We also wish to thank Dr Chao Li and Dr Xiuji Li for their help and suggestions. We wish to thank Yizhen Wei for her suggestions for language editing.

FUNDING

This work was partially supported by grants from the National Key R&D Program of China (2022ZD0116003 to C.L.), CAMS Innovation Fund for Medical Sciences (2020-I2M-C&T-A-001, 2021-I2M-1-003 to C.L.; 2021-I2M-C&T-B-011 to X.P.; 2021-1-I2M-056 to Z.W.), National Natural Scientific Foundation of China (U1611263 to R.F.), Capital's Funds for Health Improvement and Research (CFH 2020-2-4018 to C.L.), Beijing Natural Science Foundation (L202035 to C.L.) and Capital Exemplary Research Wards Project (BCRW202001 to C.L.). The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

AUTHORS' CONTRIBUTIONS

LM. C., F. R., P. X., and YB. W. designed the study. P. X., ZL. L., BC. Z. and XS. Z. drafted the manuscript. P. X., XS. Z., Z. Lv., CP. W., H. Zheng, LJ. S., N. Y., H. C., L. Y., H. Zhou, JH. Z., and Z. H. extracted the research data. ZL. L., P. X., and BC. Z. performed the statistical analysis. ZL. L., BC. Z., BY. Z., F. R., P. X., and YB. W. designed the diagnostic program. WG. Z., BY. Z. and BC. Z. supported the software deployment. HY. C. and Y. W. performed the human tests. LM. C., XM. L., Y. Q., and F. Z. supervised the study and critically reviewed the manuscript. All authors approved the final manuscript.

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared upon reasonable request to the corresponding author.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflicts of interest.

REFERENCES

1. Yang CW, Harris D, Luyckx VA *et al*. Global case studies for chronic kidney disease/end-stage kidney disease care. *Kidney Int Suppl (2011)* 2020;**10**:e24–48. https://doi.org/10.1016/j.kisu.2019.11.010
2. Lv JC, Zhang LX. Prevalence and disease burden of chronic kidney disease. *Adv Exp Med Biol* 2019;**1165**:3–15. https://doi.org/10.1007/978-981-13-8871-2_1
3. Xie J, Chen N. Primary glomerulonephritis in mainland China: an overview. *Contrib Nephrol* 2013;**181**:1–11. https://doi.org/10.1159/000348642
4. Pan X, Xu J, Ren H *et al*. Changing spectrum of biopsy-proven primary glomerular diseases over the past 15 years: a single-center study in China. *Contrib Nephrol* 2013;**181**:22–30. https://doi.org/10.1159/000348638
5. Yang Y, Zhang Z, Zhuo L *et al*. The spectrum of biopsy-proven glomerular disease in China: a systematic review. *Chin Med J (Engl)* 2018;**131**:731–5. https://doi.org/10.4103/0366-6999.226906
6. Beck LJ, Bonegio RG, Lambeau G *et al*. M-type phospholipase A2 receptor as target antigen in idiopathic membranous

nephropathy. *N Engl J Med* 2009;**361**:11–21. https://doi.org/10.1056/NEJMoa0810457
7. He P, Yu X, Zha Y *et al*. Microhematuria enhances the risks of relapse and renal progression in primary membranous nephropathy. *Front Med (Lausanne)* 2021;**8**:704830. https://doi.org/10.3389/fmed.2021.704830
8. Couser WG. Primary membranous nephropathy. *Clin J Am Soc Nephrol* 2017;**12**:983–97. https://doi.org/10.2215/CJN.11761116
9. Lees GE, Cianciolo RE, Clubb FJ. Renal biopsy and pathologic evaluation of glomerular disease. *Top Companion Anim Med* 2011;**26**:143–53. https://doi.org/10.1053/j.tcam.2011.04.006
10. Walker PD. The renal biopsy. *Arch Pathol Lab Med* 2009;**133**:181–8. https://doi.org/10.5858/133.2.181
11. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;**61**:85–117. https://doi.org/10.1016/j.neunet.2014.09.003
12. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;**521**:436–44. https://doi.org/10.1038/nature14539
13. Becker JU, Mayerich D, Padmanabhan M *et al*. Artificial intelligence and machine learning in nephropathology. *Kidney Int* 2020;**98**:65–75. https://doi.org/10.1016/j.kint.2020.02.027
14. Niel O, Bastard P. Artificial Intelligence in nephrology: core concepts, clinical applications, and perspectives. *Am J Kidney Dis* 2019;**74**:803–10. https://doi.org/10.1053/j.ajkd.2019.05.020
15. Kannan S, Morgan LA, Liang B *et al*. Segmentation of glomeruli within trichrome images using deep learning. *Kidney Int Rep* 2019;**4**:955–62. https://doi.org/10.1016/j.ekir.2019.04.008
16. Kim YG, Choi G, Go H *et al*. A fully automated system using a convolutional neural network to predict renal allograft rejection: extra-validation with giga-pixel immunostained slides. *Sci Rep* 2019;**9**:5123. https://doi.org/10.1038/s41598-019-41479-5
17. Hermsen M, de Bel T, den Boer M *et al*. Deep learning-based histopathologic assessment of kidney tissue. *J Am Soc Nephrol* 2019;**30**:1968–79. https://doi.org/10.1681/ASN.2019020144
18. Ginley B, Lutnick B, Jen KY *et al*. Computational segmentation and classification of diabetic glomerulosclerosis. *J Am Soc Nephrol* 2019;**30**:1953–67. https://doi.org/10.1681/ASN.2018121259
19. Kolachalama VB, Singh P, Lin CQ *et al*. Association of pathological fibrosis with renal survival using deep neural networks. *Kidney Int Rep* 2018;**3**:464–75. https://doi.org/10.1016/j.ekir.2017.11.002
20. Simon O, Yacoub R, Jain S *et al*. Multi-radial LBP features as a tool for rapid glomerular detection and assessment in whole slide histopathology images. *Sci Rep* 2018;**8**:2032. https://doi.org/10.1038/s41598-018-20453-7
21. Bukowy JD, Dayton A, Cloutier D *et al*. Region-based convolutional neural nets for localization of glomeruli in trichrome-stained whole kidney sections. *J Am Soc Nephrol* 2018;**29**:2081–8. https://doi.org/10.1681/ASN.2017111210
22. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;**60**:84–90. https://doi.org/10.1145/3065386
23. Ligabue G, Pollastri F, Fontana F *et al*. Evaluation of the classification accuracy of the kidney biopsy direct immunofluorescence through convolutional neural networks. *Clin J Am Soc Nephrol* 2020;**15**:1445–54. https://doi.org/10.2215/CJN.03210320
24. Zhang L, Li M, Wu Y *et al*. Classification of renal biopsy direct immunofluorescence image using multiple attention convolutional neural network. *Comput Methods Programs Biomed* 2022;**214**:106532. https://doi.org/10.1016/j.cmpb.2021.106532
25. Pan S, Fu Y, Chen P *et al*. Multi-task learning-based immunofluorescence classification of kidney disease. *Int J Environ Res Public Health* 2021;**18**:10798. https://doi.org/10.3390/ijerph182010798
26. Szegedy C, Vanhoucke V, Ioffe S *et al*. Rethinking the inception architecture for computer vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016:2818–26.
27. Turner N, Lameire N Goldsmith DJ *et al*., eds. *Oxford Textbook of Clinical Nephrology*. 4th ed. Oxford, United Kingdom: Oxford University Press, 2016:540–42.
28. Wang WX, Hu CY. Value of immunofluorescence-mediated detection of Ig, C1q, C3, and FRA for the identification and diagnosis of atypical membranous nephropathy. *Eur Rev Med Pharmacol Sci* 2017;**21**:5415–9.
29. Huang G, Liu Z, Van Der Maaten L *et al*. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017:4700–8.
30. He K, Zhang X, Ren S *et al*. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:770–8.
31. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations* 2015:1–14.
32. Zhang L, Long J, Jiang W *et al*. Trends in chronic kidney disease in China. *N Engl J Med* 2016;**375**:905–6. https://doi.org/10.1056/NEJMc1602469
33. Fu J, Zheng H, Mei T. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017:4438–46.
34. Dong HR, Wang YY, Cheng XH *et al*. Retrospective study of phospholipase A2 receptor and IgG subclasses in glomerular deposits in Chinese patients with membranous nephropathy. *PLoS One* 2016;**11**:e156263.
35. Esteva A, Robicquet A, Ramsundar B *et al*. A guide to deep learning in healthcare. *Nat Med* 2019;**25**:24–9. https://doi.org/10.1038/s41591-018-0316-z
36. Zhao K, Tang YJJ, Zhang T *et al*. DGDI: a dataset for detecting glomeruli on renal direct immunofluorescence. *2018 Digital Image Computing: Techniques and Applications (DICTA)*, Canberra, ACT, Australia, 2018:1–7.
37. Jennette JC, D'Agati VD, Olson JL *et al*. *Heptinstall's Pathology of the Kidney*. Seventh Edition. Philadelphia: Wolters Kluwer, 2015.
38. Morgenstern Y, Schmidt F, Fleming RW. A dataset for evaluating one-shot categorization of novel object classes. *Data Brief* 2020;**29**:105302. https://doi.org/10.1016/j.dib.2020.105302

Received: 28.11.2022; Editorial decision: 31.5.2023