Deep Learning-Based Model Significantly Improves Diagnostic Performance for Assessing Renal Histopathology in Lupus Glomerulonephritis

Luping Shen[a]   Wenyi Sun[a]   Qixiang Zhang[a]   Mengru Wei[a]   Huanke Xu[a]
Xuan Luo[b]   Guangji Wang[a]   Fang Zhou[a]

[a]Key Laboratory of Drug Metabolism and Pharmacokinetics, State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing, China; [b]School of Pharmacy, China Pharmaceutical University, Nanjing, China

Abstract

Background: Assessment of glomerular lesions and structures plays an essential role in understanding the pathological diagnosis of glomerulonephritis and prognostic evaluation of many kidney diseases. Renal pathophysiological assessment requires novel high-throughput tools to conduct quantitative, unbiased, and reproducible analyses representing a central readout. Deep learning may be an effective tool for glomerulonephritis pathological analysis. ***Methods:*** We developed a murine renal pathological system (MRPS) model to objectify the pathological evaluation via the deep learning method on whole-slide image (WSI) segmentation and feature extraction. A convolutional neural network model was used for accurate segmentation of glomeruli and glomerular cells of periodic acid-Schiff-stained kidney tissue from healthy and lupus nephritis mice. To achieve a quantitative evaluation, we subsequently filtered five independent predictors as image biomarkers from all features and developed a formula for the scoring model. ***Results:*** Perimeter, shape factor, minimum internal diameter, minimum caliper diameter, and number of objects were identified as independent predictors and were included in the establishment of the MRPS. The MRPS showed a positive correlation with renal score ($r = 0.480$, $p < 0.001$) and obtained great diagnostic performance in discriminating different score bands (Obuchowski index, 0.842 [95% confidence interval: 0.759, 0.925]), with an area under the curve of 0.78–0.98, sensitivity of 58–93%, specificity of 72–100%, and accuracy of 74–94%. ***Conclusion:*** Our MRPS for quantitative assessment of renal WSIs from MRL/lpr lupus nephritis mice enables accurate histopathological analyses with high reproducibility, which may serve as a useful tool for glomerulonephritis diagnosis and prognosis evaluation.

© 2022 The Author(s).
Published by S. Karger AG, Basel

Introduction

Glomerular lesions are a common phenomenon in the development and progression of many kidney diseases. They play an essential role in lupus nephritis (LN) assessment. LN is a chronic inflammatory kidney disease that occurs in 40–60% of patients with systemic lupus erythe-

Luping Shen and Wenyi Sun share co-first authorship.

Correspondence to:
Guangji Wang, guangjiwang@hotmail.com
Fang Zhou, zf1113@163.com

matosus (SLE). It is a major cause of morbidity and mortality in SLE [1]. LN is associated with a wide spectrum of kidney lesions, characterized mainly by glomerular involvement and interstitial lesions. Up to three out of 10 individuals with LN develop end-stage kidney disease and kidney failure within 15 years of their diagnosis [2]. Hence, renal pathology, especially the assessment of glomerular lesions and structures, provides an essential reference for diagnosis, treatment instructions, and prognostic evaluation of SLE.

Renal histopathology is considered the gold-standard assessment of human kidney disease. In the 2003 International Society of Nephrology/Renal Pathology Society (ISN/RPS) classification [3] – the most widely used classification in clinical practice – LN is histologically classified into six distinct classes representing different severities of renal involvement. Analysis of data from histology can provide important insights into disease pathogenesis and potential treatment strategies. However, despite the unequivocal value of histopathological findings in the diagnosis of human LN, the pathological evaluation requires the participation of experienced pathologists, is time-consuming, and is prone to human error and misjudgment. To diminish manual quantitative evaluation, deep learning (also known as deep structured learning) technology has been applied to digital whole-slide image (WSI) segmentation [4–6]. Jayapandian et al. [5] developed deep learning networks to segment histological structures on kidney biopsy and nephrectomy WSIs stained with hematoxylin & eosin, periodic acid-Schiff (PAS), silver, and trichrome. Ginley et al. [7] applied convolutional neural network (CNN) models in diabetic nephropathy glomeruli, while Zeng et al. [8] identified glomerular lesions in patients with immunoglobulin A nephropathy. Although deep learning is increasingly employed in kidney pathology, there has been no reported application in LN.

Currently, the mechanisms underlying LN pathogenesis are unclear. Animal models, used as a tool of choice for basic research into exploring the pathogenesis and finding effective therapeutic strategies against LN, are mainly lupus-prone mice. Multiple SLE murine models of different genetic backgrounds, such as (NZB/NZW) F1, MRL/lpr, B6/lpr, C3H/gld/gld, and BXSB mice [9, 10], have diverse pathological mechanisms. Among them, MRL/lpr mice serve as a model to study human LN as they exhibit manifestations similar to human disease, including glomerulonephritis, vasculitis, and arthritis. Nevertheless, the scoring scale of MRL/lpr mouse renal histo-

pathology has varied among the studies. Despite that, most of the studies reported in the literature have used a scale of 0–4 according to the scoring system of Passwell et al. [11] as previously described. Others have used a scale of 0–3 [12–14] or 0–5 [8], and increments were also different (0.5 or 1). Hence, there has been no unified quantitative assessment for lupus murine renal histopathology to obtain objective data.

The deep learning model for digital WSI analysis could eliminate or at least reduce the variations in murine renal histopathology assessment results observed among pathologists. Therefore, we aimed to develop a quantitative murine renal pathological system (MRPS) based on deep learning multiclass segmentation of MRL/lpr mouse kidney PAS-stained histology to provide a more accurate and reproducible quantitative renal scoring method.

Materials and Methods

Mouse Models

Female MRL/MpJ-Fas^lpr (MRL/lpr) mice were purchased from Shanghai Slack Laboratory Animal Co., Ltd. (Shanghai, China) at 4–5 weeks of age. They were raised at the experimental animal center of China Pharmaceutical University until 15 weeks after the onset of LN. The mice were sacrificed 15–17 weeks after the onset of LN. Female BALB/c mice, weighing 25–30 g each, were purchased from the Shanghai Branch of Beijing Vitonglihua Laboratory Animal Technology Co., Ltd. All of the mice were housed in an individual ventilation cage (IVC) system with an irradiated diet, and cages/bedding/environmental enrichment were autoclaved as per the standards. The serum and urine biochemical data of MRL/lpr mice and BALB/c mice are shown in Table 1. All animal experiments were approved by the Committee of Experimental Animal Administration of the China Pharmaceutical University.

Specimen Preparation

Mouse kidneys were fixed in Bouin solution fixative at room temperature, protected from light for 2–3 days, processed by standard procedures through graded concentrations of alcohol and xylene, and embedded in paraffin and sectioned at 2–3 μm. The sections were stained with H & E, PAS, and Masson trichrome (MASSON) in accordance with the manufacturer's instructions. All histopathological images were captured with an Olympus VS200 scanner (Olympus Corporation) at ×40 magnification and 138 nm/pixel resolution, and then processed using Olympus OlyVIA software (version 3.2). Each WSI was assessed by three independent researchers using scales of 0–4 in increments of 1, as previously described [11]. The measurement was graded into four categories: (1) (a mild increase in mesangial cellularity in the matrix, affecting <50% of glomeruli), (2) (a moderate increase in mesangial cellularity in the matrix, affecting >50% of glomeruli, with thickening of the glomerular basement membrane), (3) (focal endocapillary hypercellularity [affecting <50% of glomeruli], with obliteration of capillary lumen including a substantial increase in the thickness and irregularity of the glomerular basement mem-

Shen/Sun/Zhang/Wei/Xu/Luo/Wang/Zhou

Table 1. Characteristics of mice in the two sets

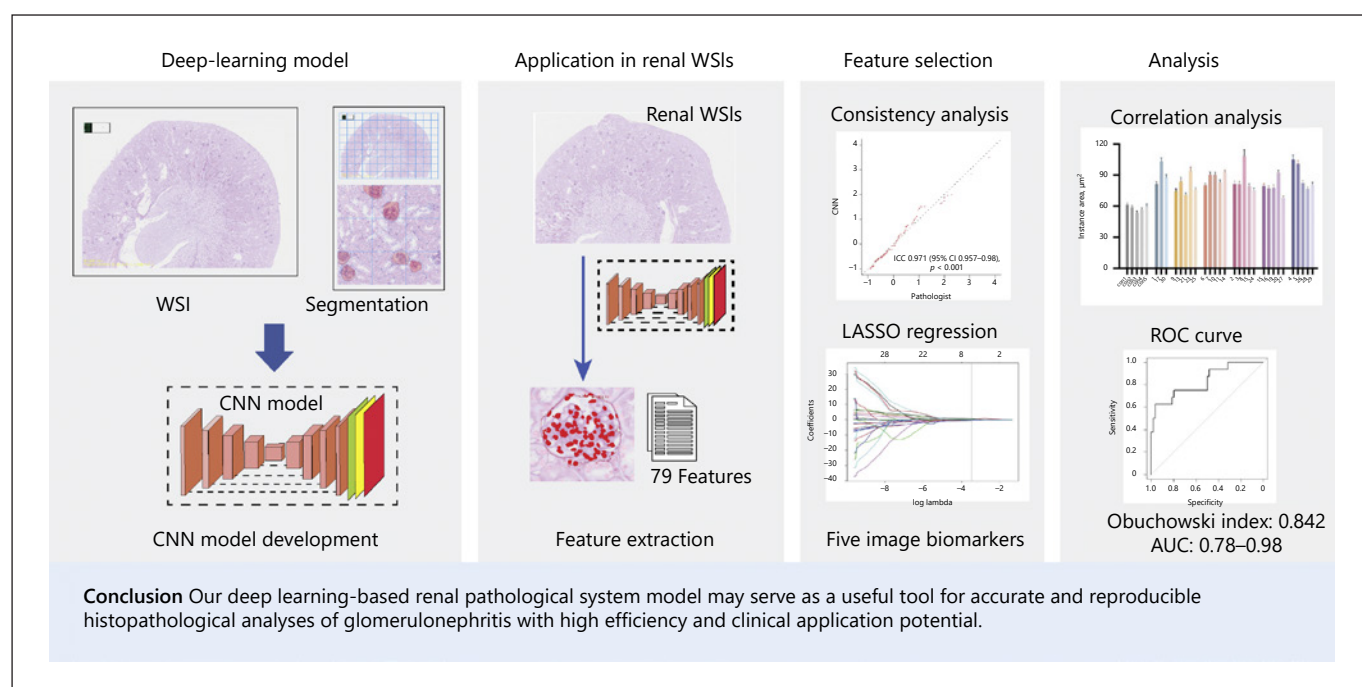|	Training set (n = 129)	Validation set (n = 70)	p value
Weight, g	37.98±3.31	37.70±4.24	0.18
Urine protein, µg/mL	10,988.24±5,902.34	11,217.78±6,406.31	0.80
Urinary creatinine, µmol/L	1,569.28±563.74	1,560.93±635.05	0.89
Serum anti-dsDNA	20.46±10.84	19.24±11.96	0.47
Albumin/creatinine, mg/µmol	7.54±4.56	6.81±4.44	0.22
Hair loss score	0.93±0.94	0.90±1.01	0.87
Skin scores	1.64±0.75	1.60±0.70	0.68
Histopathology score			
S0	9	9	
S1	39	20	
S2	18	15	0.23
S3	31	10	
S4	32	16	

Fig. 1. Workflow of all necessary steps in this study. A CNN model was trained on ImageNet dataset for image segmentation and quantification of parameters. Features were extracted using the CNN model. Selection of the majority of features was conducted by the LASSO regression, and independent factors were presented and integrated into a prediction model. The performance of the established model was evaluated by ROC analysis. LASSO, least absolute shrinkage and selection operator; ROC, receiver operator characteristic.

brane), (4) (diffuse endocapillary hypercellularity [affecting >50% of glomeruli], segmental necrosis, crescents and hyalinized end-stage glomeruli). The healthy BALB/c mice scored 0. Cases of disagreement were carefully reviewed and discussed by three senior pathologists to reach a consensus, and this consensus assignment was used as the ground truth. We selected slides showing different histopathological scores for the deep learning training model, with at least three WSIs from each score group. For developing the his-

tological assessment model, the WSIs (n = 199 in total) were split into the training set (n = 129) and the validation set (n = 70).

Training, Testing, and Assessment of the Neural Network
The whole procedure is shown in Figure 1. First, a pathologist marked glomerular areas by labeling along the outer margin of the glomerular capsule and all cell nuclei in the glomeruli (online suppl. Fig. 1; see www.karger.com/doi/10.1159/000524880 for all on-

Fig. 2. Deep learning-based model for recognition and segmentation on WSIs of murine kidney PAS staining. **a** The original WSI was split up into a multitude of single images. **b** Representative PAS pictures and corresponding segmentation predictions generated by the CNN, marked as regions of interest. **c** The mask of CNN produced annotations.
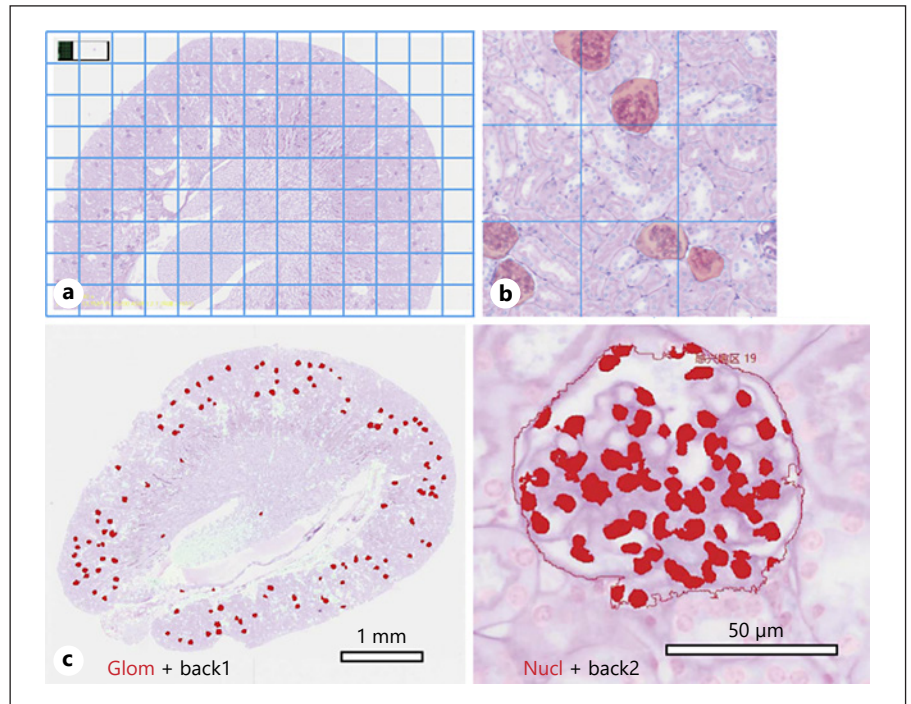
line suppl. material). Inspired by the study conducted by Bouteldja et al. [4], renal tissues (only the cortex) were grouped into four classes, including Glomeruli (Glom), Background (Back1), Nuclei in the glomeruli (Nucl), and Nuclei outside the glomeruli (Back2). As shown in online supplementary Figure 1a, the cyan parts are the identified glomeruli areas. In online supplementary Figure 1b, all renal nuclei were marked in magenta. Perform measurement in the region of glomeruli areas for identifying intrinsic glomerular cells (online suppl. Fig. 1c). These annotations were then checked by three experienced senior pathologists. We used Olympus cellSens software (Olympus Corporation, version 3.2) to train a neural network of glomerular and nuclear segmentation on a desktop workstation with an Nvidia Quadro 16 GB GPU. We employed deep learning models on the basis of the U-Net architecture using code-free deep learning with the Olympus deep-learning platform.

Due to the large size of the WSI files, which precluded directly inputting entire WSIs into the model, the original WSI was split up into a multitude of single images that were subsequently processed using the CNN (Fig. 2a). For each single image, we first located the positions of glomeruli, marked as regions of interest on the image (Fig. 2b). The training was run for a total of 25,000 iterations. The total number of glomeruli recognition training epochs was set as 50,000 to ensure convergence, and the intersection over union (IoU) was 0.77. The nuclei recognition training epoch was 1,000 times, and its IoU was 0.87. We only retained the model at the epoch of overall minimum validation loss. As the epoch of training increased, the loss decreased to less than 0.2 (online suppl. Fig. 2). The consistency analysis was performed on the evaluation set by assessing the concordance between the deep-learning model and pathologist annotation using the intraclass correlation coefficient (ICC) [15]. The agreement in deep-learning estimation be-

tween the CNN model and pathologist annotation was evaluated using the Bland-Altman plot.

Histological Feature Extraction and Selection

According to the manufacturer's instructions, we applied the deep-learning model on WSIs of MRL/lpr renal PAS staining for extraction of histopathological features. A two-step procedure was followed to select significant features. First, the correlation between each feature and the pathological score was evaluated by using the Kendall correlation coefficient to remove weakly correlated features. Features with correlation coefficients of less than 0.15 were eliminated. Then, feature selection was performed by using the least absolute shrinkage and selection operator (LASSO) logistic regression algorithm and random forest (RF) [16]. LASSO logistic regression was performed with penalty parameter tuning conducted by 10-fold cross-validation between S0–S2 and S3–S4. RF selected the most significant features according to the mean decrease accuracy (MDA) and the mean decrease Gini (MDG). Features with nonzero coefficients in LASSO logistic regression, whose importance was evaluated as being within the top 15 both by MDA and MDG, were identified as independently related to pathological score.

Statistical Analysis

Categorical and continuous variables were compared by using the χ^2 test and Student's t test, respectively. Performance of the models for renal pathology score was evaluated by receiver operating characteristics (ROC) curve analysis [17], area under the curve (AUC) value, and the Obuchowski index – a multinomial version of ROC curve analysis adapted for ordinal references such as histological renal score. The Obuchowski index is a weighted average of the AUCs obtained for all possible pairs of score bands to be
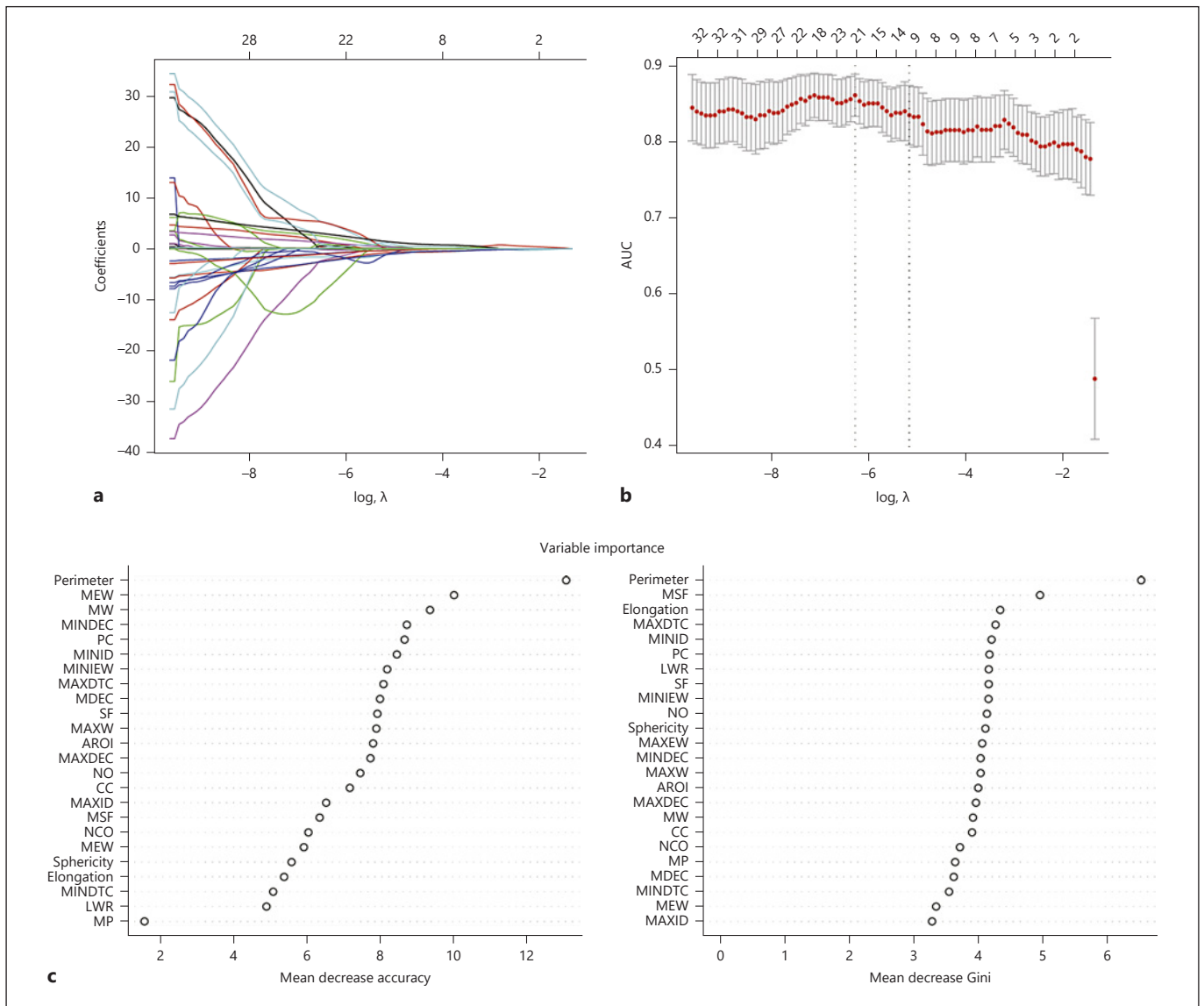
Fig. 3. Selection of features using the LASSO regression and RF. **a** Optimal λ value was determined via the LAS-SO model using 10-fold cross-validation via minimum criteria. The binomial deviance curves were plotted versus log(λ). Dotted vertical lines were drawn at the optimal values by using the minimum criteria and one standard error of the minimum criteria (the 1 – standard error criteria). The optimal λ value of 0.0018 was chosen. **b** LAS-SO coefficient profiles of the 35 selected features are presented. **c** RF ranked the importance of the initially selected features according to the MDA (left) and the MDG index (right).

differentiated. It estimates the probability that a test will correctly rank two randomly chosen patients with different renal scores. The optimal thresholds of the models were determined using the ROC analysis by maximizing the Youden index. A Delong nonparametric approach was used to compare the AUC values [18]. Statistical analysis was performed using R statistical software (version 4.1.1, www.r-project.org). A two-sided p value <0.05 was indicative of a statistically significant difference.

Results

Sample Characteristics

The baseline characteristics of all of the mice are shown in Table 1. A total of 199 mouse kidney WSIs were randomized into two sets. Both the training and the validation sets were composed of the MRL/lpr mice and BALB/c
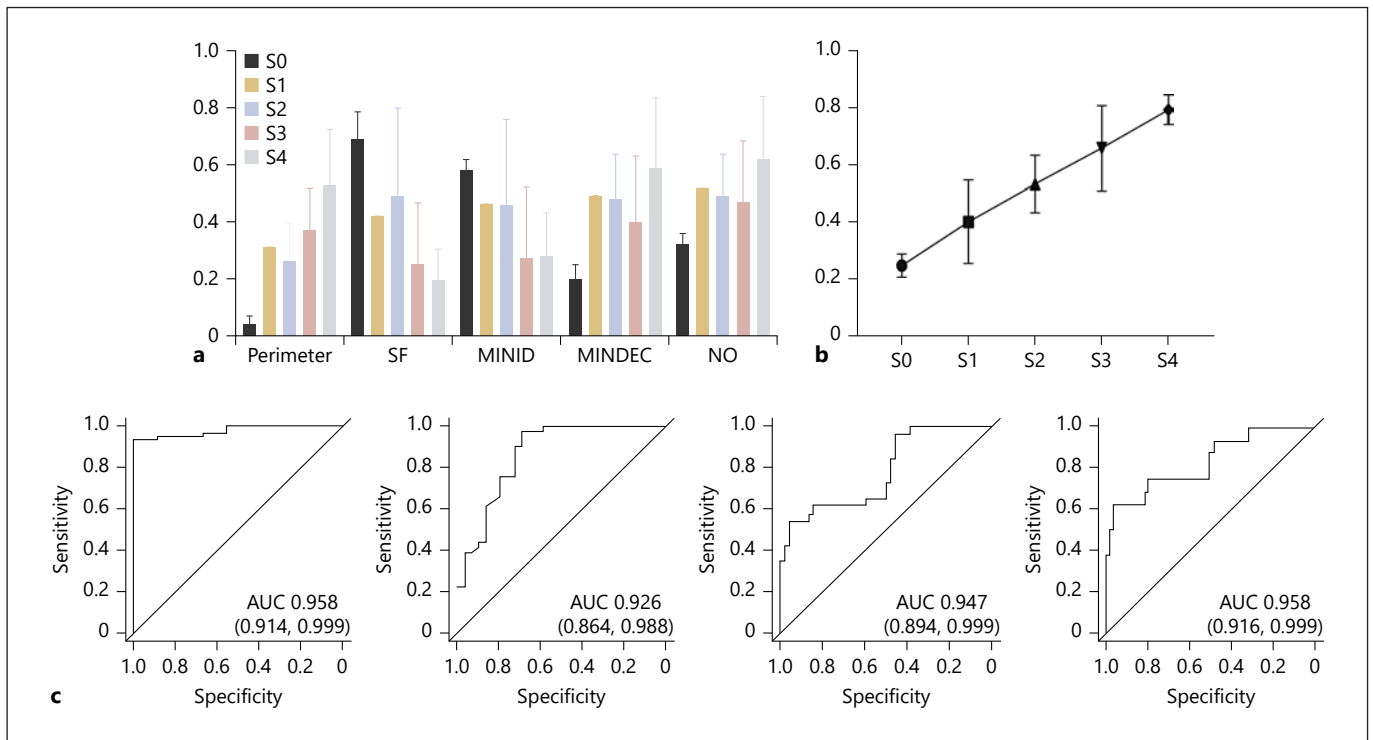
Fig. 4. Independent variables histogram and diagnostic performance of MRPS in the validation cohort. **a** Histograms of independent variables demonstrate significant correlations between the variables and the renal score ($p < 0.01$ for all). Coefficients of perimeter, SF, MINID, MINDEC, and NO were 0.465, −0.352, −0.209, 0.205, and 0.200, respectively. **b** The results calculated by MRPS in the validation cohort. Significant correlation between the results and the renal score was found ($r = 0.480$, $p < 0.001$). **c** ROC curves of MRPS for the discrimination of S0 versus S1–S4, S0–S1 versus S2–S4, S0–S2 versus S3–S4, and S0–S3 versus S4.

mice. There were no differences in terms of serum and urine biochemical parameters between the two sets. In the training set, the hair loss score was 0.93 ± 0.94, and the skin score was 1.64 ± 0.75; in the validation set, the hair loss score was 0.90 ± 1.01, and the skin score was 1.60 ± 0.70. No significant difference was found in any of these measurements. The rates of LN were 93.0% (120 of 129) and 87.1% (61 of 70) in the training and the validation sets, respectively. No significant difference was found between the two cohorts (χ^2, $p = 0.23$).

Feature Extraction with the CNN Model

The newly trained deep-learning model was then utilized to develop a novel method to access the glomeruli and identify intrinsic glomerular cells (Fig. 2c). To evaluate the consistency and the output of our CNN model, we selected two features out of 79 features in total to conduct manual measurements (online suppl. Fig. 3). Strong concordance was observed in the perimeter and counted values between the CNN model and the pathologist's measurements. A high agreement occurred between the pathologist and the CNN perimeter (ICC = 0.937; 95% CI, 0.957–0.98) or the count value (ICC = 0.994; 95% CI, 0.991–0.996). The Bland-Altman plot showed good agreement between the CNN model and the pathologist's measurements (online suppl. Fig. 3b, d). The mean difference between CNN and pathologist readouts was −0.01 in parameters (95% CI, −0.29 to 0.26; online suppl. Fig. 2b) and 0 in counts (95% CI, −0.16 to 0.16; online suppl. Fig. 2d). We applied the CNN model on renal PAS-stained WSIs and achieved 79 features on each slide for further analysis, including 38 glomerular features and 41 cellular features.

Feature Selection

Among all of the features, 35 features with significant correlations to the renal score were identified. All abbreviation of these 35 features used throughout the paper can be found in online supplementary Table 1. As illustrated in Figure 3a, 20 features with nonzero coefficients were

selected by the LASSO logistic regression. Perimeter, shape factor (SF), minimum internal diameter (MINID), minimum diameter (with an external caliper) (MINDEC), and number of objects (NO) were finally considered as independent predictors by MDA and MDG in the RF model. All of the independent predictors significantly correlated with the scoring system ($p < 0.001$ for all, Fig. 3c).

Development and Validation of the Scoring Model

A prediction model (MRPS) integrated the perimeter, SF, MINID, MINDEC, and NO. The equation for MRPS derived from the training cohort was as follows:

$$\text{MRPS} = \frac{e^a}{e^a + 1},$$

where $a = 0.103 \times \text{Perimeter} - 2.254 \times \text{SF} - 0.460 \times \text{MINID} + 1.350 \times \text{MINDEC} + 1.354 \times \text{NO}$, in which e^a is the exponential function, and a is the exponent.

The MRPS showed a positive correlation with the renal score ($r = 0.480$, $p < 0.001$; Fig. 4b). In the training cohort (Table 2), AUCs of MRPS for discriminating S0 versus S1–S4, S0–S1 versus S2–S4, S0–S2 versus S3–S4, and S0–S3 versus S4 were 0.979 (95% CI, 0.956–0.999), 0.794 (95% CI, 0.715–0.874), 0.811 (95% CI, 0.739–0.884), and 0.858 (95% CI, 0.778–0.938), respectively (Fig. 4c). The Obuchowski index was 0.841 (95% CI, 0.759–0.923).

Table 3 summarizes the diagnostic performance of MRPS in the validation cohort. MRPS had great diagnostic performance for aiding in diagnosis of S0 versus S1–S4 (0.978; 95% CI, 0.949–0.999), S0–S1 versus S2–S4 (0.864; 95% CI, 0.772–0.956), S0–S2 versus S3–S4 (0.781; 95% CI, 0.667–0.895), and S0–S3 versus S4 (0.831; 95% CI, 0.706–0.956). By using threshold values for MRPS determined in the training cohort, MRPS had a sensitivity range of 58–93%, a specificity range of 72–100%, and an accuracy range of 74–94% in discriminating S0 versus S1–S4, S0–S1 versus S2–S4, S0–S2 versus S3–S4, and S0–S3 versus S4 in the validation cohort (Table 4).

Discussion

Glomerular lesions are frequent in the pathogenesis of renal disease. They play an essential role in LN assuagement. In the 2003 ISN/RPS system for classifying patients with LN, glomerular lesions are vital for the classification. Our murine lupus model simulates human SLE, with all kinds of glomerular lesions. Thus, our MRPS model supports the possibility of generalizing to other glomerular diseases.

Table 2. Diagnostic performance of MRPS in the development cohort

Parameter	Model
S0 versus S1–S4	
AUC	0.979 (0.956, 0.999)
Threshold	0.322
Sensitivity, %	95.0 (89.4, 98.1)
Specificity, %	99.9 (66.4, 99.9)
Accuracy, %	95.3 (91.7, 99.0)
S0–S1 versus S2–S4	
AUC	0.794 (0.715, 0.874)
Threshold	0.509
Sensitivity, %	60.5 (49.0, 71.2)
Specificity, %	83.3 (69.8, 92.5)
Accuracy, %	69.0 (60.9, 77.1)
S0–S2 versus S3–S4	
AUC	0.811 (0.739, 0.884)
Threshold	0.612
Sensitivity, %	71.4 (58.7, 82.1)
Specificity, %	81.8 (70.4, 90.2)
Accuracy, %	76.7 (69.4, 84.1)
S0–S3 versus S4	
AUC	0.858 (0.778, 0.938)
Threshold	0.665
Sensitivity, %	78.1 (60.0, 90.7)
Specificity, %	82.5 (73.4, 89.4)
Accuracy, %	81.4 (74.6, 88.2)
Obuchowski index	0.841 (0.759, 0.923)

Data in parenthesis are 95% confidence intervals. AUC, area under the curve.

Table 3. Diagnostic performance of MRPS in the validation cohort

Parameter	Model
AUC	
S0 versus S1–S4	0.978 (0.949, 0.999)
S0–S1 versus S2–S4	0.864 (0.772, 0.956)
S0–S2 versus S3–S4	0.781 (0.667, 0.895)
S0–S3 versus S4	0.831 (0.706, 0.956)
Obuchowski index	0.842 (0.759, 0.925)

Data in parenthesis are 95% confidence intervals. AUC, area under the curve.

MRL/lpr mice are the most commonly used animal model in LN basic research, which shows a full spectrum of glomerulonephritis, closely resembling the pathological manifestations of human SLE. Since the change of MRL/lpr mouse glomeruli and intrinsic glomerular cells highly correlates with the basic lesion in LN, we designed

Table 4. Diagnostic performance of MRPS in the validation cohort

Parameter	S0 versus S1–S4	S0–S1 versus S2–S4	S0–S2 versus S3–S4	S0–S3 versus S4
Threshold value	0.322	0.509	0.612	0.665
Sensitivity, %	93.4 (57/61) [87.1, 99.8]	78.0 (32/41) [64.8, 91.3]	57.7 (15/26) [37.3, 78.0]	62.5 (10/16) [35.9, 89.1]
Specificity, %	100 (9/9) [100, 100]	72.4 (21/29) [55.1, 89.7]	84.1 (37/44) [72.8, 95.3]	90.7 (49/54) [82.8, 98.7]
Accuracy, %	94.3 (66/70) [88.7, 99.9]	75.7 (53/70) [65.4, 86.0]	74.3 (52/70) [63.8, 84.8]	84.3 (59/70) [75.5, 93.0]

Data in parenthesis are numerator/denominator and data in brackets are 95% confidence intervals.

MRPS to make the scoring process more objective. We chose the following five parameters as the evaluation indicators of our model: perimeter, SF, MINID, MINDEC, and NO. Perimeter, MINID, and MINDEC are used to describe the glomerular size, while SF measures the glomerular shape. The number of objects refers to the counts of intrinsic glomerular cell numbers. All these glomerular morphometric values are associated with the severity of nephritis [19]. Lupus glomerular lesions involve pathological changes such as thickening of the glomerular basement membrane, expansion of the mesangial area, proliferation of mesangial cells, and sclerosis in focal segmental glomeruli. An alteration in any of these pathological signs could show abnormalities in the five indicators. For instance, hyperplasia and inflammatory cell infiltration would increase the number of intrinsic glomerular cells. The increased collagen deposition and collapsing glomerulopathy have been associated with glomerular shape. In brief, the five variables integrated into the MRPS model could be image biomarkers of glomerulonephritis.

Renal pathology of LN involves proliferation of endothelial and mesangial cells, the basement membrane with wire-loop capillaries reminiscent, glomerular crescent, renal interstitial nephritis, as well as vasculitis. Due to the complexity, it is difficult to identify, label, and accurately measure each value. To solve the fatiguing manual analysis, we employed deep learning methods to identify glomerular features. Deep learning is a subset of intelligence that applies computer algorithms to meaningful representations of raw data through multiple layers of abstraction. It enabled the effective process of WSI segmentation. With success in recent studies that applied CNNs to reveal prognostic biomarkers directly from digitalized WSIs [4, 8], our work further efficiently facilitated the extraction of pathological information for unambiguous outcome indicators of renal scores.

Given that the MRPS model has the potential to be applied to various glomerular diseases, as well as provide a standard morphometric analysis on renal tissues, it is a good tool for preclinical and clinical studies. In both preclinical and clinical practice, histopathologic evaluations are often performed manually, which is both time-consuming and not seldom poorly reproducible, particularly if not performed by experienced senior researchers and experts. When applying our model, pathologists are no longer needed.

Our method also showed simplicity and specificity for clinical application. It is possible to achieve promising recognition accuracy in different renal lesions with relatively low additional annotation effort by experts. This might allow rapid adaptation of the method to samples from various laboratories and translation to additional models and pathologies. It is an important prerequisite for high-throughput and reproducible pathology analyses and will be essential to reduce the workload while at the same time increasing the quantitative precision in experimental and potentially also clinical histopathology. For both basic research and clinical-pathological assessment, there is also a strong demand for the analysis of high-throughput WSI analysis. Our study provided a new idea for renal pathology based on deep learning and is beneficial to preclinical nephritis models that also might have the potential to be used on other sample types, such as clinical FFPE samples, which also might be clinically applied in the future.

Nowadays, there are an increasing number of tools to generate the segmentation of the entire WSIs [20], such as HALO software [21] and the PyTorch platform [22, 23]. However, they require either GPU-based hardware or programming skills for researchers, which limits the promotion and application. Furthermore, with so many ways to assess, extract, and compute a large number of image features, the pathology reports are still too descriptive to quantify lesions even precisely scored. Our MRPS system combined the code-free platform and scoring formula to address these challenges. The CNN model solved the segmentation of WSIs, and it was able to be easily used on a personal computer. The scoring part quantified glo-

merular lesions to contribute to overcoming the observer's potential bias.

Our study should be considered in light of some implicit limitations. First, we carried out a single mouse model. It needs to be generalized and extended to other murine models of lupus in the future. Besides, our WSI data set was collected at different times with different batches but not from different laboratories. We hope to conduct the cross-laboratory validation in the future. Furthermore, we only focused on glomerulonephritis/ glomeruli lesions and did not further discriminate some glomerular morphometric evaluations, such as endocapillary hypercellularity, mesangial matrix expansion, and tubulointerstitial lesions. In addition, we did not conduct a comprehensive assessment of all renal lesions. Finally, our MRPS cannot be automated. It requires two steps that involve WSI's quantitative segmentation and scoring model. This performance still needs to be improved to be more time-saving.

In conclusion, this study presents MRPS, which provides a digital indicator for murine renal pathology. Our system can provide a computational approach for the lupus mouse histopathology scoring based on the deep learning recognition of glomeruli and intrinsic glomerular cells. It has an excellent universality and high applicability for basic scientists and could be used directly for assisting the pathological diagnosis of LN.

Acknowledgments

Statement of Ethics

All animal care and experimental procedures were conducted according to the National Research Council's Guidelines for the Care and Use of Laboratory Animals and were approved by the SPF Animal Laboratory of China Pharmaceutical University (Animal Authorization Reference Number: SYXK2016-0011).

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

Author Contributions

Fang Zhou and Guangji Wang conceived the strategy for this study. Luping Shen and Fang Zhou designed all the experiments. Luping Shen, Wenyi Sun, and Qixiang Zhang prepared the murine sample. Wenyi Sun, Mengru Wei, and Huanke Xu assisted with pathology staining of the slides and acquisition of images. Luping Shen and Xun Luo assisted with the statistical analysis. Fang Zhou and Luping Shen wrote the manuscript with comments from all the authors.

Data Availability Statement

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

References

1 Sung SJ, Fu SM. Interactions among glomerulus infiltrating macrophages and intrinsic cells via cytokines in chronic lupus glomerulonephritis. J Autoimmun. 2020;106:102331.

2 Wong T, Goral S. Lupus nephritis and kidney transplantation: where are we today? Adv Chronic Kidney Dis. 2019;26(5):313–22.

3 Weening JJ, D'Agati VD, Schwartz MM, Seshan SV, Alpers CE, Appel GB, et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. Kidney Int. 2004;65(2):521–30.

4 Bouteldja N, Klinkhammer BM, Bülow RD, Droste P, Otten SW, Freifrau von Stillfried S, et al. Deep learning-based segmentation and quantification in experimental kidney histopathology. J Am Soc Nephrol. 2021;32(1):52–68.

5 Jayapandian CP, Chen Y, Janowczyk AR, Palmer MB, Cassol CA, Sekulic M, et al. Development and evaluation of deep learning-based segmentation of histologic structures in the kidney cortex with multiple histologic stains. Kidney Int. 2021;99(1):86–101.

6 Hermsen M, de Bel T, den Boer M, Steenbergen EJ, Kers J, Florquin S, et al. Deep learning-based histopathologic assessment of kidney tissue. J Am Soc Nephrol. 2019;30(10):1968–79.

7 Ginley B, Lutnick B, Jen KY, Fogo AB, Jain S, Rosenberg A, et al. Computational segmentation and classification of diabetic glomerulosclerosis. J Am Soc Nephrol. 2019;30(10):1953–67.

8 Zeng C, Nan Y, Xu F, Lei Q, Li F, Chen T, et al. Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning. J Pathol. 2020;252(1): 53–64.

9 Stoll ML, Gavalchin J. Systemic lupus erythematosus-messages from experimental models. Rheumatology. 2000;39(1):18–27.

10 Perry D, Sang A, Yin Y, Zheng YY, Morel L. Murine models of systemic lupus erythematosus. J Biomed Biotechnol. 2011;2011: 271694.

11 Passwell J, Schreiner GF, Nonaka M, Beuscher HU, Colten HR. Local extrahepatic expression of complement genes C3, factor B, C2, and C4 is increased in murine lupus nephritis. J Clin Invest. 1988;82(5):1676–84.

12 Tshilela KA, Ikeuchi H, Matsumoto T, Kuroiwa T, Sakurai N, Sakairi T, et al. Glomerular cytokine expression in murine lupus nephritis. Clin Exp Nephrol. 2016;20(1):23–9.

13 Kikawada E, Lenda DM, Kelley VR. IL-12 deficiency in MRL-Fas(lpr) mice delays nephritis and intrarenal IFN-gamma expression, and diminishes systemic pathology. J Immunol. 2003;170(7):3915–25.

14 Ma H, Liu C, Shi B, Zhang Z, Feng R, Guo M, et al. Mesenchymal stem cells control complement C5 activation by factor H in lupus nephritis. EBioMedicine. 2018;32:21–30.

15 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–8.

16 Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol. 2016;34(18):2157–64.

17 Lambert J, Halfon P, Penaranda G, Bedossa P, Cacoub P, Carrat F. How to measure the diagnostic accuracy of noninvasive liver fibrosis indices: the area under the ROC curve revisited. Clin Chem. 2008;54(8):1372–8.

18 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–45.

19 Kiberd BA. The functional and structural changes of the glomerulus throughout the course of murine lupus nephritis. J Am Soc Nephrol. 1992;3(4):930–9.

20 Huo Y, Deng R, Liu Q, Fogo AB, Yang H. AI applications in renal pathology. Kidney Int. 2021;99(6):1309–20.

21 Horai Y, Mizukawa M, Nishina H, Nishikawa S, Ono Y, Takemoto K, et al. Quantification of histopathological findings using a novel image analysis platform. J Toxicol Pathol. 2019; 32(4):319–27.

22 Athavale AM, Hart PD, Itteera M, Cimbaluk D, Patel T, Alabkaa A, et al. Development and validation of a deep learning model to quantify interstitial fibrosis and tubular atrophy from kidney ultrasonography images. JAMA Netw Open. 2021;4(5):e2111176.

23 Kuo CC, Chang CM, Liu KT, Lin WK, Chiang HY, Chung CW, et al. Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. NPJ Digit Med. 2019;2: 29.