BMC Nephrology

RESEARCH

Open Access

Vision transformer introduces a new vitality to the classification of renal pathology

Ji Zhang[2†], Jia Dan Lu[1†], Bo Chen[2], ShuFang Pan[1], LingWei Jin[1], Yu Zheng[1] and Min Pan[1*]

Abstract

Recent advancements in computer vision within the field of artificial intelligence (AI) have made significant inroads into the medical domain. However, the application of AI for classifying renal pathology remains challenging due to the subtle variations in multiple renal pathological classifications. Vision Transformers (ViT), an adaptation of the Transformer model for image recognition, have demonstrated superior capabilities in capturing global features and providing greater explainability. In our study, we developed a ViT model using a diverse set of stained renal histopathology images to evaluate its effectiveness in classifying renal pathology. A total of 1861 whole slide images (WSI) stained with HE, MASSON, PAS, and PASM were collected from 635 patients. Renal tissue images were then extracted, tiled, and categorized into 14 classes on the basis of renal pathology. We employed the classic ViT model from the Timm library, utilizing images sized 384 × 384 pixels with 16 × 16 pixel patches, to train the classification model. A comparative analysis was conducted to evaluate the performance of the ViT model against traditional convolutional neural network (CNN) models. The results indicated that the ViT model demonstrated superior recognition ability (accuracy: 0.96–0.99). Furthermore, we visualized the identification process of the ViT models to investigate potentially significant pathological ultrastructures. Our study demonstrated that ViT models outperformed CNN models in accurately classifying renal pathology. Additionally, ViT models are able to focus on specific, significant structures within renal histopathology, which could be crucial for identifying novel and meaningful pathological features in the diagnosis and treatment of renal disease.

Keywords Artificial Intelligence, Convolutional neural networks, Vision transformers, Renal pathology, Whole-slide imaging

Introduction

The diagnosis of kidney diseases relies on the assessment of medical images, such as computed tomography and histopathological images. However, accurately discerning the wide array of changes in renal pathology remains challenging even for experienced renal pathologists. Various traditional techniques have been developed for kidney segmentation across different imaging modalities to aid in diagnosing kidney diseases [1–4]. Although some research has addressed the classification of kidney pathologies, more advanced and effective methods are still needed in this field [5, 6]. Recent advancements in artificial intelligence (AI) have led to significant progress

†Ji Zhang Ph.D. and JiaDan Lu M.S. contributed equally to this work.

*Correspondence:
Min Pan
pm_1234@163.com
[1]Department of Nephrology, The Second Affiliated Hospital, Yuying Children's Hospital of Wenzhou Medical University, 109 Xueyuan Road, Wenzhou, Zhejiang, PR China
[2]Department of Nephrology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, PR China

in computer vision [7, 8] and its gradual expansion into the medical field [9]. Moreover, the recent developments in WSI technology have facilitated the integration of AI into renal pathology applications. Owing to its unparalleled learning capabilities, AI has the potential to acquire extensive medical knowledge from large-scale renal pathology data, thereby supporting clinical decision-making and reducing diagnostic and therapeutic bias in clinical practice [10]. However, significant obstacles still hinder the wider adoption of AI in renal pathology [11–13]. Currently, image recognition techniques and segmentation models have limited applicability in renal pathology. This is primarily due to subtle variations across multiple classifications of renal pathology and the relatively small size of available datasets. Achieving accurate labeling and recognition of these subtle pathological changes in the kidneys requires a substantial cohort of specialized nephropathologists and significant time investment. Renal pathology typically involves integrating results from multiple stains. The use of AI algorithms for integrating these results is still being explored. Furthermore, significant challenges persist in higher-level semantic recognition, such as directly identifying specific pathological changes in particular areas or types of pathological images, with corresponding research reports currently being sparse.

Currently, most models focus on segmenting various renal pathological structures, including glomeruli, renal tubules, renal blood vessels, and the renal-tubule interstitium. For example, Bel and Hermsen et al. developed and compared two fully convolutional networks, assessing the performance of UNet for classification tasks [12, 14, 15]. Meanwhile, Salvi and Bueno et al. employed semantic segmentation techniques to identify glomeruli and analyze their morphological changes [6, 16]. Other research has utilized object detection methods to recognize glomeruli within WSI [17, 18], and additional studies have classified pathological changes in glomeruli [19, 20]. Furthermore, advanced language analyses are being explored to connect renal pathological changes with clinical features such as kidney function, proteinuria, and prognosis [21]. All these approaches are grounded in CNN frameworks.

While the CNN is efficient in extracting general structural features, it faces challenges in capturing focal microfeatures. Transformers are currently considered state-of-the-art methods in nearly all natural language processing (NLP) benchmarks and have given rise to vision transformers (ViT) for application in visual tasks [22, 23]. Their approach differs from that of CNNs as they employ a self-attention mechanism to recognize image content and are more sensitive to local pixel changes over a wide range. The precise delineation of features by ViT is more likely to be extracted for the recognition of kidney pathology, thereby infusing a renewed vigor into the identification of kidney pathology images.

Our study employed four conventional staining pathological images of kidney tissue to train corresponding image classification models based on ResNet50 and ViT models. Additionally, in order to investigate the focus areas of the obtained ViT model for pathological diagnosis and the potential pathological significance of these areas, we visualized the ViT model via the Deep Taylor Decomposition principle [24], and developed a random mask method for further validation.

Methods

Experimental design and participants

The experimental process design is illustrated in Fig. 1. We conducted a retrospective review of patients who underwent ultrasound-guided percutaneous renal biopsy (PRB) at the Department of Nephrology, Second Affiliated Hospital of Wenzhou Medical University, from January 2013 to April 2020. Pathological diagnosis was performed using four routine staining techniques: hematoxylin and eosin (H&E), Masson's trichrome (Masson), Periodic Acid-Schiff (PAS), and Periodic Schiff-Methenamine (PASM). WSIs were obtained by scanning the stained renal tissue slides with a KF-PRO whole-slide scanner manufactured by KFBIO. A dataset comprising patients with at least two WSI exhibiting different staining patterns was constructed.

Preprocessing of digital pathological sections

The WSIs of the included patients were converted to RGB bitmap images at maximum magnification, which acquired at ×20 accessory magnification, resulting in images with a resolution of 0.5 μM per pixel. Afterwards, the entire image was tiled into 384 * 384 pixels tiles. To account for the presence of various non-renal tissue regions, such as fat, muscle, and background, we applied a combination of color thresholding and edge detection, followed by manual screening, to identify renal tissue regions. Tiles that overlapped with renal regions by more than 90% were selected for further analysis (Fig. 1a-d). In order to address the imbalance in the number of images across different pathological diagnosis types, we employed a rotation augmentation technique to equalize the number of images across groups. Subsequently, the acquired tiles set was randomly partitioned into training and validation sets at an 8:2 ratio.

Model construction and model training

We employed the ResNet50 and ViT models, both of which were constructed using the timm library [25]. To ensure a precise evaluation of whether the VIT model outperforms ResNet50 in kidney pathology image recognition tasks, we maintain the original model structure

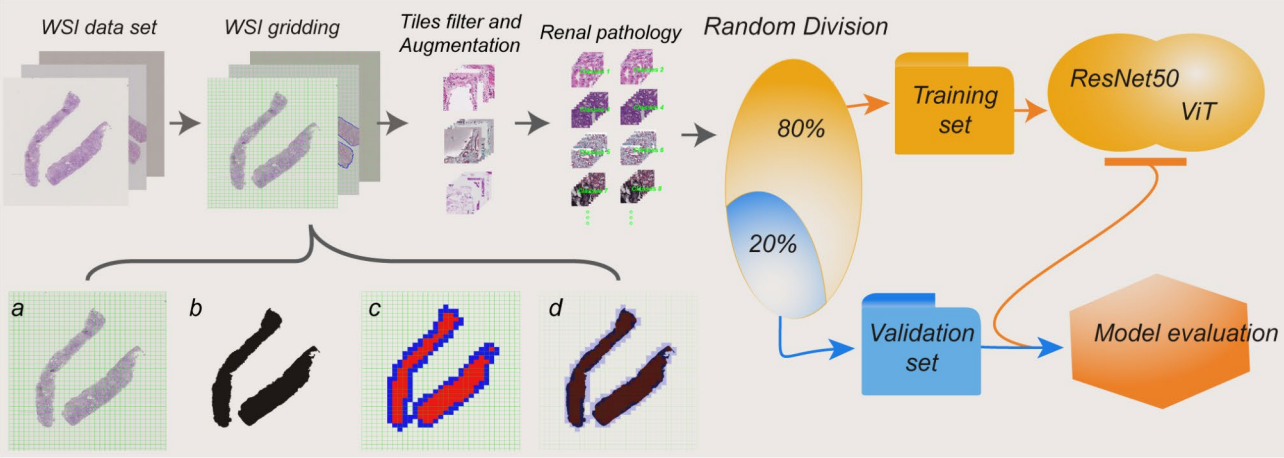

Fig. 1 Overview of the period pipeline. The collected WSI images are used to construct a WSI dataset, and then the WSI is transformed into a grid format. Kidney tissue regions are identified, and a mask image is built with grid images to obtain kidney tissue of interest. The obtained grid images are used to construct an image set and are randomly divided into a training set and a validation set. The ResNet50 and ViT models are trained on training set images. Finally, the model performance is verified via a validation set. Figure a. Gridded pathological image, where b results from identified kidney tissue regions and a mask is constructed. In c, the red grid shows the grid where the proportion of kidney tissue is greater than 90%, and the blue grid shows the grid where the proportion of kidney tissue is less than 90%. Figure d is a fusion image of Figures b and c, which more obviously shows the position of the grid in the renal tissue

without any modifications. In this configuration, the image size is set to 384×384 pixels, divided into patches of 16×16 pixels each. The model has 14 different types, a depth of 12 layers, and 12 attention heads. The multilayer perceptron (MLP) ratio is set to 4, indicating the ratio of the hidden layer size to the input layer size within the MLP block. The embedding dimension is 768, resulting in a total parameter count of 86,859,496. Considering the efficacy of importing pretraining parameters in enhancing model performance and reducing training cycles, all the models were initialized with pretraining parameters. We trained separate classification models for each of the four staining methods, conducting 100 epochs per model. The timm library was employed to construct an optimizer. Although hybrid loss functions have been used in some recent works [26–28], the cross-entropy loss function is computationally efficient and commonly used. Therefore, it has been used in our experimental wırks.

Performance evaluation

Our validation dataset was used to evaluate the identification performance of the ResNet50 and ViT classification models across four different stainings. Additionally, we employed the F1 score and Receiver Operating Characteristic (ROC) curves to assess variations in recognition performance among different renal pathological types.

Visualization and Random Mask Verification of the ViT Model

For the trained ViT model, we aim to identify the specific pathological structures that significantly influence its classification performance to detect critical pathological changes. To visualize and analyze the impact of different regions on the model's prediction results, we employed Chefer et al.'s ViT visualization method, which generates heatmaps [24]. Inspired by the mask method in the NLP model, we hypothesize that significant changes in classification occur when a specific area in an image is filled with blank space. These findings suggest that the identified area may play a crucial role in identifying the pathological type of renal disease, indicating its importance as an essential pathological change for that specific type. To validate this hypothesis, we developed a random mask method in which 256 random 64 * 64 boxes were generated for each validation image. The pixels within these boxes were replaced with zeros and subsequently used for image classification. Any misclassifications occurring within the box areas were considered indicative of their significance.

Results

Images incorporation and Dataset construction

We enrolled a total of 635 patients, 46.4% of whom were female with an average age of 43.4 years. A comprehensive collection of 1861 high-quality renal pathology WSIs were obtained, comprising HE-stained (n=633), MASSON-stained (n=635), PAS-stained (n=633), and PASM-stained (n=628) samples. The dataset processing is illustrated in Fig. 1. In brief, the images represented 14

distinct pathological types, each meticulously reviewed by PM and CB for diagnostic accuracy. Since this study focused solely on classification training, cases with multiple pathological diagnoses were assigned to the most prominent category after thorough discussions within the research team. The distribution of the pathological profiles of the included patients is shown in Fig. 2.

Different performances of the ViT model and ResNet50 model.

Our findings demonstrate that all four staining models exhibited high recognition performance (Accuracy range: 0.89–0.99), with HE displaying the lowest recognition ability and PASM demonstrating the most robust recognition capability, as depicted in Fig. 3. Compared with ResNet50 (Accuracy: 0.89–0.95), the ViT model showcased superior vital recognition ability (Accuracy: 0.96–0.99). Furthermore, it is evident from F1 scores and ROC curves that ViT outperformed ResNet50 in recognizing various pathological types. However, the model's recognition ability for IgAN and MN was relatively weak, which could be attributed to a greater number of cases.

Model interpretation

The model interpretation is illustrated in Fig. 4. A heatmap was used to visualize the attention points of the ViT models, revealing that after training, the primary renal pathological information captured by the models was concentrated in the glomerulus, tubules, and tubulointerstitium. Conversely, there was no significant enhancement in the blank area, indicating that the models were able to identify essential pathological features. Furthermore, our findings from random mask plot and Deep Taylor Decomposition principle results demonstrated substantial overlap, providing further validation for the potential diagnostic significance of these renal structures. Additionally, we highlight that changes around the tubulointerstitial area may constitute an important component of renal pathology that is often overlooked by pathologists. We observed a concentration of self-attention around the renal tubules in the LN, potentially linked to IgG deposition. These subtle pixel changes are imperceptible to the naked eye. Additionally, PASM staining for IgA nephropathy revealed an accumulation of hot spots in the renal tubules and mesangial area, possibly associated with immune complex deposition in the mesangial region.

Discussion

Traditionally, the diagnosis of renal pathology usually relies on clinical manifestations and results from different types of staining via light microscopy and additional fluorescence and electron microscopy. However, computer vision can identify some fine distinctions of structures or color changes that are difficult to distinguish with the naked eye, providing a new approach for accurate diagnosis and improvement of renal pathological diagnosis. We used ResNet50 and the latest ViT framework to train corresponding recognition models for the four types of renal pathology stainings, and the results showed that they can effectively distinguish the different renal pathologies. We found that the ViT models are superior to the

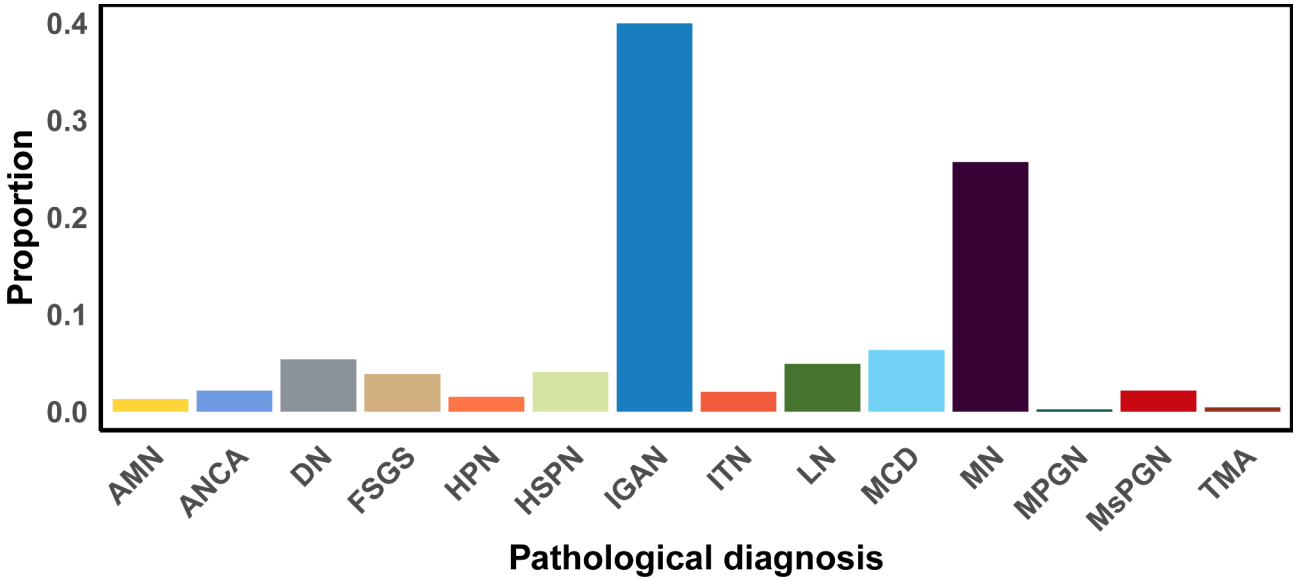

Fig. 2 Renal pathological distribution of the included patients. AMN: amyloidosis nephropathy; ANCA: ANCA-associated glomerulonephritis; DN: diabetic nephropathy; FSGS: focal segmental glomerulosclerosis; HPN: hypertension-attributed nephropathy; HSPN: Henoch–Schönlein purpura nephritis; IGAN: Henoch–Schönlein purpura nephritis; ITN: tubulointerstitial nephritis; LN: lupus nephritis; MCD: minimal change disease; MN: membranous nephropathy; MPGN: membranoproliferative glomerulonephritis; MsPGN: mesangial proliferative glomerulonephritis; TMA: renal thrombotic microangiopathy

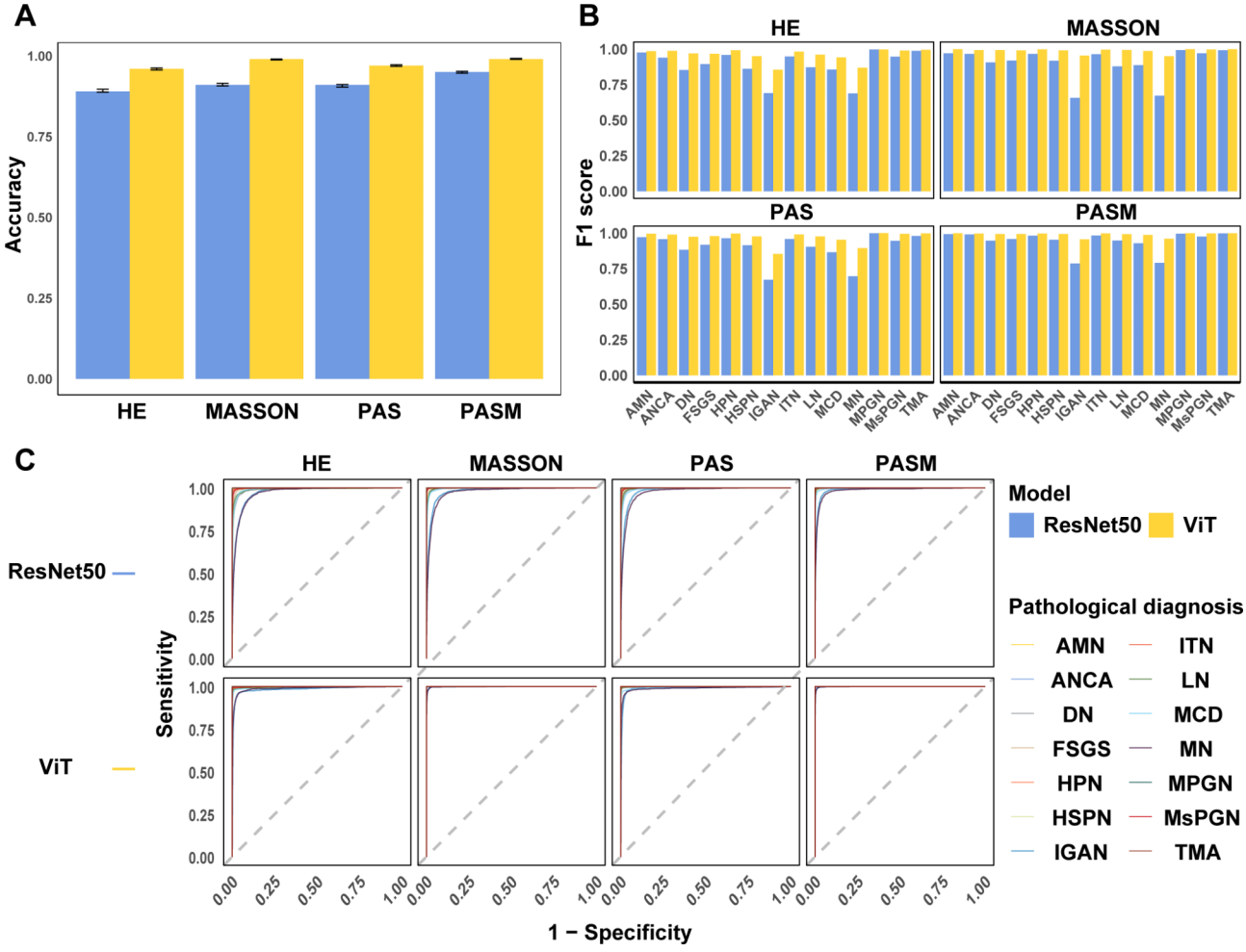

Fig. 3 The difference in prediction accuracy between ResNet50 and ViT models in the validation set. Figure A shows the accuracy scores of the ResNet50 and ViT models. ViT model has a higher accuracy score in all pathological types. Figure B shows the F1 scores for the ResNet50 and ViT models in different pathological classes. The models trained using PASM-stained pathological images have higher F1 scores. Figure C uses receiver operating characteristic, showing that all curves are approximately symmetrical, the curve area under the ViT models are larger, and the Masson and PASM were the most performance models

ResNet50 models' identification performance for the four staining methods. Furthermore, the ViT model outperforms the ResNet50 framework in terms of recognition performance for different pathological types.

Previous AI research on renal pathology has focused mainly on recognizing structures involved in renal pathology, including glomeruli, tubules, renal vessels, and some inflammatory cells, which involve image segmentation and semantic recognition [21, 29–34]. Although significant progress has been made in these areas of research, more advanced pathological diagnosis are still lacking. Renal pathological diagnosis involves various aspects of the disease and must be combined with various types of renal pathological staining. Our study revealed that the models trained with the PASM and Masson staining methods can achieve greater accuracy. It is speculated that PASM staining and Masson staining provide a more comprehensive color distribution and corresponding pathological tissues, which makes the pathological images more informative and conducive to the realization of renal pathological diagnosis [5].

The architecture of the ViT model differs from that of traditional deep convolutional models in that it recognizes images through a self-attention mechanism. In contrast, extracting fine-grained features from images is more better than using a CNN [35, 36]. Although traditional deep convolutional models have a firmer grasp of the overall features of the image, local features are smoothed, which is not conducive to feature recognition of renal pathology. This may also explain why the performance of the ViT model is better than that of the ResNet50 model.

We prioritize the assessment of distinct pathological structures to accurately identify renal pathology, which helps in detecting significant renal pathological changes and potentially uncovering critical alterations

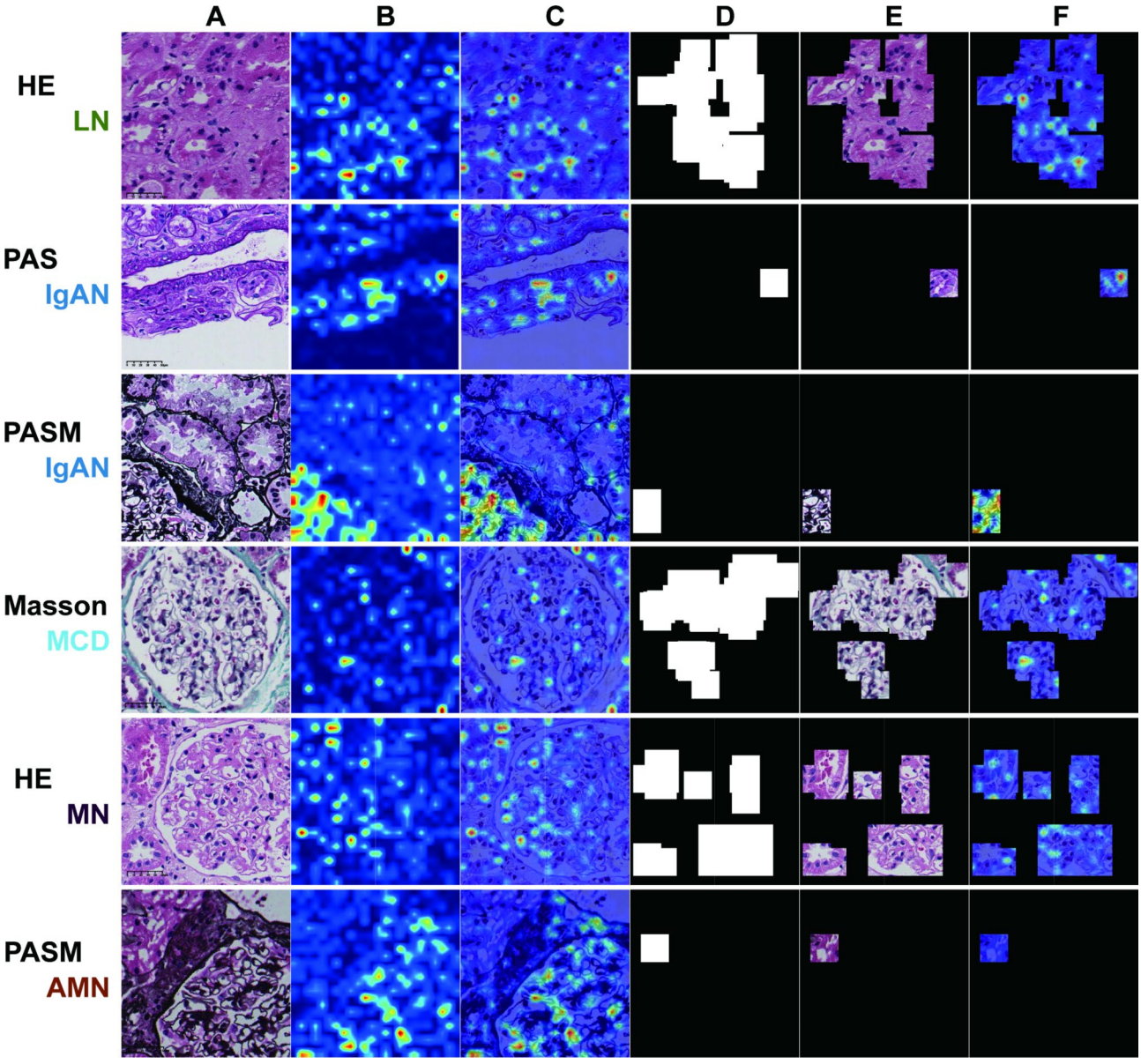

Fig. 4 ViT model heatmap visualization and random mask results. Column A depicts the original image, while column B showcases a heatmap generated using the Deep Taylor Decomposition method, with pixels displayed in a gradient from red to blue indicating high to low self-attention. Column C illustrates the merged image of A and B. Column D represents the mask graph, with white squares indicating significant random mask blocks in the image. Column E displays the merged image of A and D, while column F presents the merged image of C and D

The visualization of the ViT model heatmap and random mask results is presented. Column A depicts the original image, while column B showcases a heatmap generated using the Deep Taylor Decomposition principle method, with pixels displayed in a gradient of red to blue indicating elevated self-attention to low. Column C illustrates the merge graph of A and B, column D represents the mask graph, with white squares denoting significant random mask blocks in the image. Column E displays the merge graph of A and D, while column F presents the merge graph of C and D

that may be imperceptible to the naked eye. By visualizing the weights assigned to different renal pathological structures in our ViT model recognition process, we observed a strong alignment between the key information recognized by our model and the aspects we focused on daily. For instance, IgA nephropathy is predominantly localized in the mesangial region, and the model also provides valuable insights into areas where structural assessment is challenging. However, certain focal points may not align with conventional observations. The first image in Column C of Fig. 4 illustrates that the interstitial area surrounding the renal tubules is also a region of interest in IgA nephropathy. While the interstitial area is not a site for IgA deposition, tubulointerstitial changes are intricately linked to sclerosis in the mesangial area and are closely associated with

the prognosis of IgA nephropathy [37–40]. Therefore, certain characteristic interstitial changes may represent specific structural features of IgA nephropathy that have yet to be acknowledged. In the context of LN, the model highlights peritubular changes that may not be readily apparent through H&E staining alone. The incorporation of fluorescence in SLE demonstrates the rapid deposition of immune complexes surrounding the renal tubules, potentially contributing to the model's validation. We further use the random mask method to validate the heatmap results, which shows that the hot spot area is masked while causing model recognition errors. The mask method confirms that the heatmap results have significant reference values. However, inconsistent findings have also been reported. For instance, in the case of AMN, increased signal intensity is noted in the glomerular region, while the random mask area corresponds to the renal tubules. Both regions are significant for detecting abnormal light chain deposition, indicating that further investigation may be necessary.

Some limitations of our study are that although the recognition performance of our model was excellent in the validation set, we did not validate it on external data. There is a risk of overfitting, which can be reduced by further expanding the sample size. Second, there are often multiple pathological changes in renal pathological classification, and we identify only the most significant pathological changes, leading to poor performance and insufficient generalization performance. In addition, because the resolution of our images after scanning is still low, it will also lead to the blurring of some renal pathological features and affect the model recognition performance.

Although direct diagnosis of renal pathology by light microscopy and four kinds of staining is challenging, our study shows that it is possible to directly identify renal pathological types by light microscopy, which has important implications for renal pathological diagnosis. The ViT model can achieve better results than the ResNet50 model in the classification of renal pathology, indicating that the ViT model may be more suitable for recognizing renal pathology, especially for recognizing local fine particle features. Through heatmap and the random Mask method, the focus of the model was shown to be consistent with the characteristics of renal pathological changes, which may indicate that some renal pathological changes are difficult to distinguish by the naked eye, providing a new way to explore the renal pathological changes of related types of renal pathology. A similar approach can be applied to histopathological images of liver tissues. While numerous studies have focused on liver segmentation using various methods [41–43], the classification of fatty liver tissues remains insufficiently addressed. As another future work, the effectiveness of the ViT model can be compared with the effectiveness of capsule networks due to their ability of them in keeping spatial relationships of learned features and yield high performance in classifying medical images [44–46]. Hence, the ViT model holds significant potential for aiding renal pathology diagnosis in the foreseeable future.

Acknowledgements
I would like to express my gratitude to Hugo Touvron et al. for their outstanding contributions to this work. The authors thank their colleagues in the Department of Nephrology, The Second Affiliated Hospital of Wenzhou Medical University, for their support and help during this period.

Author contributions
Ji Zhang, Yu Zheng and JiaDan Lu: design, conceptualization and formal analysis of the research study, writing and editing of the manuscript; Bo Chen, LingWei Jin and ShuFang Pan: collection, mark the image, and analysis of data; Min Pan: conceptualization of the research study, funding acquisition for the study, critical review and revision of the manuscript.

Funding sources
This research was funded by the Medical Health Science and Technology Project of Zhejiang Provincial Health Commission (grant No. 2020377868), and the Wenzhou Science & Technology Bureau (grant No. Y20210150).

Data availability
The renal histopathological Images could not be made public due to ethical concerns. The codes for data generated or analysed during this study can be directed to the corresponding author.

Declarations

Statement of Ethics
All patients enrolled in the study were provided with comprehensive information regarding the potential benefits and risks associated with renal biopsy, as well as the possibility of utilizing their clinical data for future clinical research purposes. Subsequently, written informed consent was obtained from each participant prior to conducting the renal biopsy procedure. Furthermore, meticulous strategies are implemented to ensure the protection of patients' personal information while extracting clinical data and histopathology images. This protocol was reviewed and approved by the Ethics Review Committee of the Second Affiliated Hospital of Wenzhou Medical University, approval number LCKY2019-217.

Consent for publication
NA.

Competing interests
The authors declare no competing interests.

Received: 17 March 2024 / Accepted: 7 October 2024
Published online: 09 October 2024

References
1. Goceri N, Goceri E. A Neural Network Based Kidney Segmentation from MR Images. In: 2015 *IEEE 14th International Conference on Machine Learning and Applications (ICMLA): 9–11 Dec. 2015 2015*; 2015: 1195–1198.
2. Goceri E. Automatic kidney segmentation using Gaussian Mixture Model on MRI sequences. Electrical Power Systems and computers: 2011// 2011; Berlin. Heidelberg: Springer Berlin Heidelberg; 2011. pp. 23–9.
3. Islam MSB, Sumon MSI, Sarmun R, Bhuiyan EH, Chowdhury MEH. Classification and segmentation of kidney MRI images for chronic kidney disease detection. Comput Electr Eng. 2024;119:109613.
4. Inoue K, Hara Y, Nagawa K, Koyama M, Shimizu H, Matsuura K, Takahashi M, Osawa I, Inoue T, Okada H, et al. The utility of automatic segmentation of kidney MRI in chronic kidney disease using a 3D convolutional neural network. Sci Rep. 2023;13(1):17361.

5. Jiang L, Chen W, Dong B, Mei K, Zhu C, Liu J, Cai M, Yan Y, Wang G, Zuo L, et al. A deep learning-based Approach for Glomeruli Instance Segmentation from Multistained Renal Biopsy pathologic images. Am J Pathol. 2021;191(8):1431–41.
6. Salvi M, Mogetta A, Gambella A, Molinaro L, Barreca A, Papotti M, Molinari F. Automated assessment of glomerulosclerosis and tubular atrophy using deep learning. Comput Med Imaging Graph. 2021;90:101930.
7. Goceri E. Vision transformer based classification of gliomas from histopathological images. Expert Syst Appl. 2024;241:122672.
8. Goceri E. Nuclei segmentation using attention aware and adversarial networks. Neurocomputing. 2024;579:127445.
9. Cossio M. Computational pathology in renal disease: a comprehensive perspective. arXiv Preprint arXiv:221010162 2022.
10. Huo Y, Deng R, Liu Q, Fogo AB, Yang H. AI applications in renal pathology. Kidney Int. 2021;99(6):1309–20.
11. Feng C, Liu F. Artificial intelligence in renal pathology: current status and future. Biomolecules Biomed. 2023;23(2):225.
12. Hermsen M, de Bel T, Den Boer M, Steenbergen EJ, Kers J, Florquin S, Roelofs JJ, Stegall MD, Alexander MP, Smith BH. Deep learning–based histopathologic assessment of kidney tissue. J Am Soc Nephrol. 2019;30(10):1968–79.
13. Boor P. Artificial intelligence in nephropathology. Nat Rev Nephrol. 2020;16(1):4–6.
14. Bel Td, Hermsen M, Smeets B, Hilbrands LB, Laak Jvd, Litjens GJS. Automatic segmentation of histopathological slides of renal tissue using deep learning. In: *Medical Imaging*: 2018; 2018.
15. Hermsen M, Volk V, Bräsen JH, Geijs DJ, Gwinner W, Kers J, Linmans J, Schaadt NS, Schmitz J, Steenbergen EJ, et al. Quantitative assessment of inflammatory infiltrates in kidney transplant biopsies using multiplex tyramide signal amplification and deep learning. Lab Invest. 2021;101(8):970–82.
16. Bueno G, Fernández-Carrobles MdM, González-López L, Déniz-Suárez O. Glomerulosclerosis identification in whole slide images using semantic segmentation. Comput Methods Programs Biomed. 2019;184:105273.
17. Rehem JMC, dos Santos WLC, Duarte AA, de Oliveira LR, Angelo MF. Automatic glomerulus detection in renal histological images. Medical imaging 2021: Digital Pathology: 2021. SPIE; 2021. pp. 115–25.
18. Kaur G, Garg M, Gupta S, Juneja S, Rashid J, Gupta D, Shah A, Shaikh A. Automatic Identification of Glomerular in Whole-Slide Images Using a Modified UNet Model. *Diagnostics (Basel)* 2023, 13(19).
19. Weis C-A, Bindzus JN, Voigt J, Runz M, Hertjens S, Gaida MM, Popovic ZV, Porubsky S. Assessment of glomerular morphological patterns by deep learning algorithms. J Nephrol. 2022;35(2):417–27.
20. Cascarano GD, Debitonto FS, Lemma R, Brunetti A, Buongiorno D, De Feudis I, Guerriero A, Venere U, Matino S, Rocchetti MT. A neural network for glomerulus classification based on histological images of kidney biopsy. BMC Med Inf Decis Mak. 2021;21:1–14.
21. Kolachalama VB, Singh P, Lin CQ, Mun D, Belghasem ME, Henderson JM, Francis JM, Salant DJ, Chitalia VC. Association of pathological fibrosis with renal survival using deep neural networks. Kidney Int Rep. 2018;3(2):464–75.
22. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst 2017, 30.
23. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv 2020. *arXiv preprint arXiv:201011929* 2010.
24. Chefer H, Gur S, Wolf L. Transformer interpretability beyond attention visualization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 2021*; 2021: 782–791.
25. Wightman R. PyTorch image models, 2019, 10.5281/zenodo. 4414861. *GitHub Repository*https://www.githubcom/rwightman/pytorch-image-models
26. Goceri E. Polyp segmentation using a hybrid vision transformer and a hybrid loss function. J Imaging Inf Med. 2024;37(2):851–63.
27. Göçeri E. An application for automated diagnosis of facial dermatological diseases. İzmir Katip Çelebi Üniversitesi Sağlık Bilimleri Fakültesi Dergisi. 2021;6(3):91–9.
28. Goceri E. GAN based augmentation using a hybrid loss function for dermoscopy images. Artif Intell Rev. 2024;57(9):1–19.
29. Samsi S, Jarjour WN, Krishnamurthy A. Glomeruli segmentation in H&E stained tissue using perceptual organization. In: *2012 IEEE Signal Processing in Medicine and Biology Symposium (SPMB): 2012*: IEEE; 2012: 1–5.
30. Bukowy JD, Dayton A, Cloutier D, Manis AD, Staruschenko A, Lombard JH, Woods LCS, Beard DA, Cowley AW. Region-based convolutional neural nets for localization of glomeruli in trichrome-stained whole kidney sections. J Am Soc Nephrol. 2018;29(8):2081–8.
31. Marsh JN, Matlock MK, Kudose S, Liu T-C, Stappenbeck TS, Gaut JP, Swamidass SJ. Deep learning global glomerulosclerosis in transplant kidney frozen sections. IEEE Trans Med Imaging. 2018;37(12):2718–28.
32. Gadermayr M, Dombrowski A-K, Klinkhammer BM, Boor P, Merhof D. CNN cascades for segmenting sparse objects in gigapixel whole slide images. Comput Med Imaging Graph. 2019;71:40–8.
33. Gadermayr M, Gupta L, Appel V, Boor P, Klinkhammer BM, Merhof D. Generative adversarial networks for facilitating stain-independent supervised and unsupervised segmentation: a study on kidney histology. IEEE Trans Med Imaging. 2019;38(10):2293–302.
34. Kannan S, Morgan LA, Liang B, Cheung MG, Lin CQ, Mun D, Nader RG, Belghasem ME, Henderson JM, Francis JM. Segmentation of glomeruli within trichrome images using deep learning. Kidney Int Rep. 2019;4(7):955–62.
35. Attias R. Vision transformers for analyzing High-Resolution Pathology images. Ecole Polytechnique Fédérale de Lausanne; 2023.
36. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S. An image is worth 16x16 words: transformers for image recognition at scale. arXiv Preprint arXiv:201011929 2020.
37. Zhang J, Zhou R, Mi Y, Liu Z, Huang B, Guo R, Wang P, Quan S, Zhou Y. Role of Human Mesangial-Tubular Crosstalk in Secretory IgA-Induced IgA Nephropathy. Kidney Blood Press Res. 2021;46(3):286–97.
38. Wu J, Hu Z, Wang Y, Hu D, Yang Q, Li Y, Dai W, Zhu F, Yang J, Wang M, et al. Severe glomerular C3 deposition indicates severe renal lesions and a poor prognosis in patients with immunoglobulin A nephropathy. Histopathology. 2021;78(6):882–95.
39. Dai H, Hu W, Lin L, Wang L, Chen J, He Y. Tubular decoy receptor 2 as a predictor of prognosis in patients with immunoglobulin A nephropathy. Clin Kidney J. 2021;14(5):1458–68.
40. Sheng L, Zhuang S. New insights into the role and mechanism of partial epithelial-mesenchymal transition in kidney fibrosis. Front Physiol. 2020;11:569322.
41. Göçeri E, Ünlü MZ, Dicle O. A comparative performance evaluation of various approaches for liver segmentation from SPIR images. Turkish J Electr Eng Comput Sci. 2015;23(3):741–68.
42. Dura E, Domingo J, Göçeri E, Martí-Bonmatí L. A method for liver segmentation in perfusion MR images using probabilistic atlases and viscous reconstruction. Pattern Anal Appl. 2018;21(4):1083–95.
43. Goceri E, Unlu MZ, Guzelis C, Dicle O. An automatic level set based liver segmentation from MRI data sets. In: 2012 *3rd International conference on image processing theory, tools and applications (IPTA): 2012*: IEEE; 2012: 192–197.
44. Goceri E. Classification of skin cancer using adjustable and fully convolutional capsule layers. Biomed Signal Process Control. 2023;85:104949.
45. Goceri E. Analysis of capsule networks for image classification. In: *International conference on computer graphics, visualization, computer vision and image processing: 2021*; 2021: 53–60.
46. Goceri E. Capsule neural networks in classification of skin lesions. In: *International conference on computer graphics, visualization, computer vision and image processing: 2021*; 2021: 29–36.

Publisher's note