

Received 14 February 2025, accepted 24 March 2025, date of publication 28 March 2025, date of current version 8 April 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3555654

RESEARCH ARTICLE

Renal Biopsy Pathological Tissue Segmentation: A Comprehensive Review and Experimental Analysis

YAN LIU^{1,2}, TAIPING WANG^{①3}, XINXIN HE^{1,2}, WEI WANG⁴, QIANG HE^{1,2}, AND JUAN JIN^{①,2}

¹Department of Nephrology, The First Affiliated Hospital, Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), Hangzhou, Zhejiang 310000, China

²Zhejiang Key Laboratory of Research and Translation for Kidney Deficiency-Stasis-Turbidity Disease, Hangzhou, Zhejiang 310000, China

³Hangzhou Medipath Intelligent Technology Company Ltd., Hangzhou, Zhejiang 310000, China

⁴Department of Ultrasound Diagnosis, The First Affiliated Hospital, Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), Hangzhou, Zhejiang 310000, China

Corresponding authors: Qiang He (qianghe1973@126.com) and Juan Jin (20233004@zcmu.edu.cn)

This work was supported by the “Pioneer” and “Leading Goose” Research and Development Program of Zhejiang under Grant 2023C03075.

ABSTRACT Traditional renal biopsy pathological analysis relies heavily on subjective judgment by pathologists, which is time-consuming and susceptible to inter-observer variability, especially in segmenting glomeruli, renal tubules, and renal vessels. This study aims to develop an efficient and accurate segmentation method for renal pathological tissues to support intelligent diagnostic workflows. By integrating classic models such as Unet, U²net, and EfficientNet, along with the state-of-the-art Transformer-based SwinUnet, this research conducts a comprehensive comparison of different architectures in terms of accuracy, speed, and computational resource consumption. Experimental results show that while SwinUnet achieves an F1 score of 84.02 with an inference time of 630s, the proposed EfficientNet-b4+Unet model reaches a higher F1 score of 84.26 with a significantly reduced inference time of 420s, demonstrating its superior efficiency. A complete pipeline is proposed, encompassing data preprocessing, model training, and validation, including multi-scale feature extraction and optimized loss function configurations. Moreover, to enhance segmentation accuracy at tissue boundaries, this study introduces a novel boundary processing strategy, where boundaries are treated as a separate class and assigned higher weights. Experimental validation confirms the effectiveness of this approach. Results on datasets with PAS, PASM, H&E, and Masson staining demonstrate that the improved EfficientNet-b4+Unet model achieves the best balance between performance and speed while exhibiting strong generalization across different staining methods. This study provides a novel technical framework for precise renal pathological tissue segmentation and lays the foundation for future lesion detection and intelligent diagnostic applications.

INDEX TERMS Renal pathological tissue segmentation, medical image segmentation, convolutional neural networks, Unet.

I. INTRODUCTION

Renal biopsy is an important examination method for determining the nature and pathological types of kidney diseases, developing treatment strategies for patients, and

The associate editor coordinating the review of this manuscript and approving it for publication was Vincenzo Conti^④.

predicting the prognosis of kidney diseases. It promotes the etiological diagnosis and reveals the pathological process of kidney diseases so that patients can obtain effective targeted treatment. However, at present, histopathological evaluation of renal biopsies mainly depends on the judgment of pathologists, which is time-consuming. In addition, there were significant diagnostic differences between different

pathologists. To address these issues, this study investigates a tissue segmentation algorithm in renal biopsy to provide an efficient and accurate scheme.

Recently, deep learning methods, particularly convolutional neural networks (CNNs), have been widely used in medical image analysis and have obtained good result. As an important branch of pathology, renal biopsy pathology has unique complexity in medical image analysis. To improve diagnostic efficiency and accuracy, it is important to apply deep learning methods [1], [2], [3], [4], [5], [6], [7], [8], [9], [10] to renal biopsy pathology. Related studies have been conducted on deep learning renal pathological segmentation methods [11], [12], [13], [14]. However, several problems still need to be investigated: 1) Most of the studies [15], [16], [17], [18], [19] are on specific lesions, and the reusability of these methods is lacking. In particular, studies on renal tissue mainly focus on the glomerulus [15], [16], [17], [18], [19] and occasionally involve renal tubules [20], but seldom consider renal vessels [21], [22], [23]; 2) The existing studies use a relatively simple staining method (e.g., periodic acid-Schiff, PAS) for pathological sections [15], [20], [21], [22], [24], without considering other staining methods, such as Masson, periodic acid silver methenamine (PASM), hematoxylin and eosin (H&E), etc. [25], [26]; 3) The existing studies focus on the application level [7], [8], [17], [18], [27], and an in-depth comparison of the algorithms, such as different model structures and more detailed training techniques, are lacking [23], [28]. 4) Most existing segmentation methods struggle with precise boundary detection, particularly for renal vessels and tubules, where unclear boundaries lead to misclassification [15], [16], [17], [18], [19]. Current approaches lack explicit boundary refinement techniques to improve segmentation accuracy in complex pathological images. 5) The importance of multi-scale feature extraction [29] has been overlooked in previous studies. Proper feature extraction is essential for capturing both fine-grained tissue details and global structural information across different staining types and magnifications. A more effective multi-scale approach is needed to enhance segmentation robustness. 6) Although the latest ConvNext and Transformer-based models [29], [30], [31], [32], [33], [34], [35], [36], [37] have been extensively studied in medical image analysis, their application and systematic evaluation in renal pathology segmentation remain limited. Further research is required to validate their effectiveness in this domain.

In renal biopsy, the pathological tissue mainly includes the glomerulus, tubules, and vessels. These tissues form the basis for renal disease diagnosis; therefore, it is important to accurately segment them. In recent years, various methods, including instance segmentation and semantic segmentation techniques, have been applied to renal tissue segmentation, each with its own limitations.

For example, Bae et al. [38] proposed some instance segmentation methods in their study. The first method, based

on YOLOv8, demonstrated high speed and efficiency in renal tissue segmentation tasks. However, its segmentation accuracy was relatively low, with significant omission errors, especially in complex pathological images. The second method, using Mask R-CNN, improved segmentation accuracy compared to YOLOv8 and achieved better localization of renal structures. Nevertheless, Mask R-CNN still suffered from high omission rates and significantly slower inference speed, limiting its practicality in real-time applications.

In the field of semantic segmentation, Dimitri et al. [39] and Hara et al. [40] explored DeepLabv3 and U-Net, respectively. While U-Net showed slightly better accuracy and boundary detail processing, both methods struggled with precise boundary localization and robustness in complex backgrounds, particularly for renal vessel segmentation tasks.

With the rise of Transformer-based models in image segmentation, Feng et al. [41] compared SwinUnet and ConvWinUnet. These models significantly improved segmentation accuracy through global context modeling, but their slower inference speeds and higher computational requirements present challenges for practical applications in pathology diagnostics.

Despite the advances in transformer-based segmentation models, prior research has not comprehensively evaluated models such as ConvNext or recent variations of Swin Transformer, which may offer better trade-offs between accuracy and efficiency. This study addresses this gap by incorporating a broader range of model architectures and conducting a systematic comparison to evaluate their suitability for renal pathology segmentation.

To address these limitations, this study integrates boundary-aware processing techniques and multi-scale feature extraction to enhance segmentation accuracy. The proposed method employs a novel boundary refinement strategy that treats tissue boundaries as a separate class and assigns higher weights, improving precision in contour delineation. Additionally, multi-scale feature extraction is incorporated into the model architecture to effectively capture both global structural information and local fine-grained details, ensuring robust performance across different staining methods and imaging conditions.

Combined with the most advanced deep learning algorithm, this study proposes a segmentation algorithm for pathological renal tissue. The algorithm involves pretreatment, a model structure, a training module, and a testing module, in which the model structure is divided into backbone, neck, and head. Each component was designed and tested in this study. Specifically, 1) In the pretreatment stage, a boundary refinement technique is introduced, experimentally demonstrating improved segmentation accuracy. Additionally, the impact of different patch sizes and input resolutions on model performance is explored. 2) In the model structure, a series of classic and most advanced models, such as Unet [42], [43], [44], U²net [45], EfficientNet [46], and Transformer, were compared, and

some improvements were made. In addition, different neck and head modules were used to perform experiments on feature fusion. 3) In the training module, some training details are presented, and the effects of BN, batch size, and different loss functions on the model effect are compared. 4) In the test module, the specific operation steps are given, and some post-processing methods to improve the effectiveness of the model are also shown.

Finally, by comparing nearly 100 groups of experiments, we selected two groups of models. The model based on U²net showed strong performance, although it was slightly weak in speed, and the model based on EfficientNet-b4+Unet achieved a balance between performance and speed.

II. MATERIALS

All samples used in this study were provided by The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine) after de-identification (without patient information), and approval of the Research Ethics Committee of the Zhejiang Provincial Hospital of Chinese Medicine (Acceptance number 2024-KL-01). All biopsy samples were processed using standard light and electron microscopy techniques. Under the light microscope, formalin-fixed and paraffin-embedded tissues were cut into slices (2 mm) and stained with one of the following methods: H&E, PAS, PASM, or Masson staining. A total of 1340 samples were obtained, including 339 samples stained with H&E, 356 samples stained with PAS, and 329 samples stained with PASM. These WSIs are characterized by extremely high resolutions, typically reaching tens of thousands by tens of thousands of pixels, and feature irregular and diverse pathological tissue boundaries. These large-scale images present unique challenges for segmentation algorithms.

Each sample, including glomeruli, tubules, and vessels, was marked by professional renal pathology mapping personnel (Figure 1). The pathologist reviewed and confirmed all labels. A total of 16852 glomerulus, 411028 tubules, and 70046 renal vessels were labeled from the above 1340 whole slide images (WSIs). For convenience of description, in this study, the glomerulus is recorded as KG1, the renal tubules as KT1, and the renal vessels as KV1.

1) Data Augmentation

To address the challenges posed by large-scale WSIs and enhance the diversity of the training dataset, this study employed a variety of data augmentation techniques. Standard augmentations, including rotation, scaling, flipping, and color space transformations, were applied. In addition, a novel adaptive random patch cropping strategy was developed to better utilize the unique characteristics of pathological images. This strategy, implemented using the Palgo platform, dynamically extracts random patches from the original WSIs during each training iteration. Real-time patch cropping ensures that the training data changes with each epoch,

effectively increasing data diversity even with a limited number of samples. Furthermore, this adaptive cropping approach avoids the generation of black edges caused by rotation, preserving the integrity of the training samples. These strategies allow the model to generalize more effectively to unseen data.

2) Statistical Analysis

To ensure the representativeness of the dataset, statistical analysis was conducted to compare the distributions of target structures (glomeruli, renal tubules, and blood vessels) across training and testing datasets. An independent samples t-test was performed to assess significant differences in the distributions, confirming that the datasets are balanced and reflective of real-world pathological scenarios. This analysis provides a robust foundation for evaluating the performance of the proposed algorithms.

3) Validation Method

To evaluate the robustness and generalization ability of the proposed model, a comprehensive validation framework was implemented. A 5-fold cross-validation strategy was applied to the 100 training samples, which were divided into five subsets. In each fold, one subset was used as the test set, while the remaining four subsets served as the training set. This process ensures that all training samples are evaluated multiple times in different configurations, providing a thorough assessment of model performance across varied data splits. Additionally, an independent validation set comprising 48 samples (12 for each staining method) was used to further assess the model's generalization ability. By combining 5-fold cross-validation on the training samples with evaluation on an independent validation set, the framework offers a robust and reliable method for performance evaluation, ensuring the model is rigorously tested under diverse conditions and simulates real-world scenarios.

III. METHODS

Figure 2 shows the general procedure of the deep learning segmentation algorithm proposed in this study, which can be divided into pretreatment, model structure, training, and testing modules. The model structure consists of a backbone, neck, and head. This study analyzes each module in detail and presents an optimal scheme.

A. PRETREATMENT

WSIs of renal pathological tissue are very large and have unique characteristics; therefore, it is necessary to carry out targeted pretreatment. Specifically, it includes the following aspects.

1) Patch Images cropping

In this study, the information of the valid blocks was stored in memory, which saved the storage space required by pre-cutting and the time cost caused by

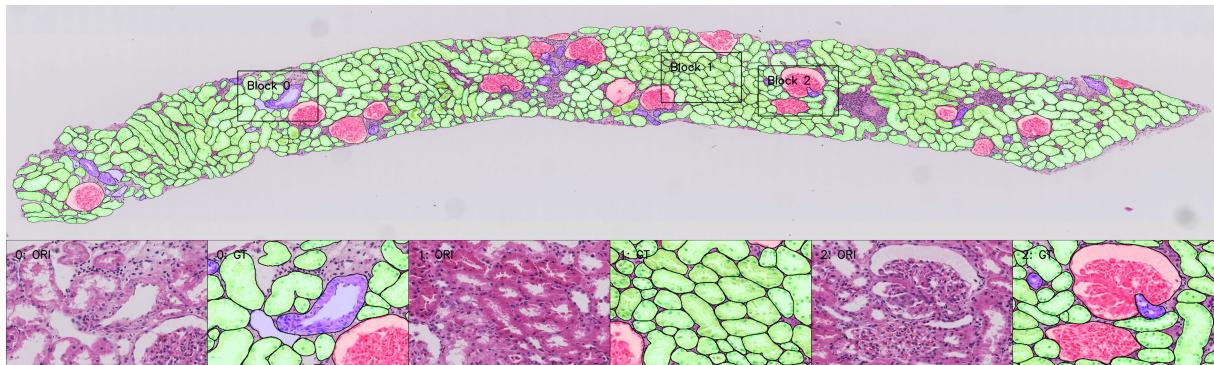


FIGURE 1. The ground truth of renal pathological tissue segmentation model. Red indicates glomerulus, green indicates tubules, and blue indicates renal vessels. And several representative blocks are selected here for detailed display. Among them, block 0 mainly shows vessels, block 1 mainly shows tubules, and block 2 mainly shows glomerulus.

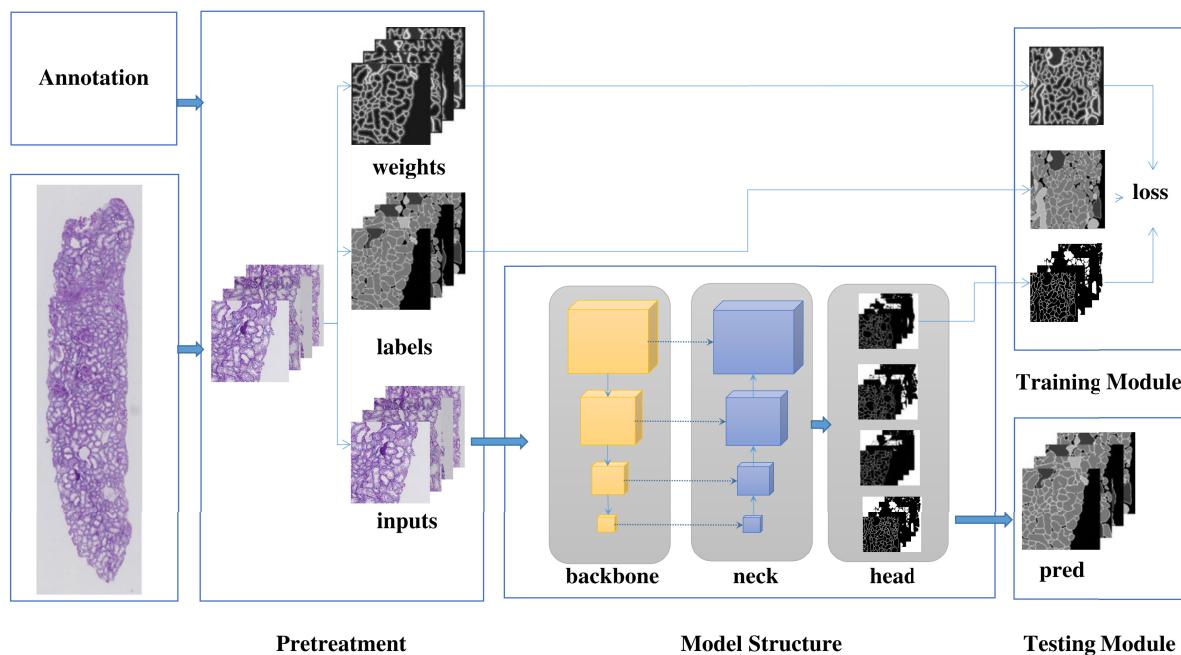


FIGURE 2. The composition of the renal pathological tissue segmentation algorithm. The algorithm can be simply divided into the following parts: pretreatment (patch images cropping, input scale, data augmentation, etc.), model structure (mainly including backbone, neck, and head), training module (loss function, optimization method), testing module (post-test processing).

repeated data reading. Meanwhile, the image selection was centered on the target to avoid marginalizing important information. Finally, the block calibration technique (filter the patch if more than 85% of the area of the ground truth were used in other patches) was used to avoid resampling of the samples.

2) Selection of image input scale

Input size has a certain impact on the efficiency and performance of the model. In this study, multi-scale comparative experiments were conducted on EfficientNet and U²net models.

3) Processing method of the target boundary

The boundary of the renal pathological tissue needs to be accurately segmented, as shown in Figure 3C, this study adopts the following approaches to deal with

the boundary. 1) The target boundary is regarded as a separate class, and a certain width is fixed on the contour of the target as the boundary. 2) Weights are added to the boundary. This study experimentally verified the effectiveness of this boundary treatment method.

4) Data augmentation

Data augmentation is performed on the patch image, including rotation, translation, cropping, blurring, noise, color contrast, and resizing [47], [48].

B. MODEL STRUCTURE

Referring to the logic of MMDetection [49], this study divides the model structure into the backbone, neck, and head. Subsequently, advanced model algorithms were employed to conduct experiments on these three components. Based on

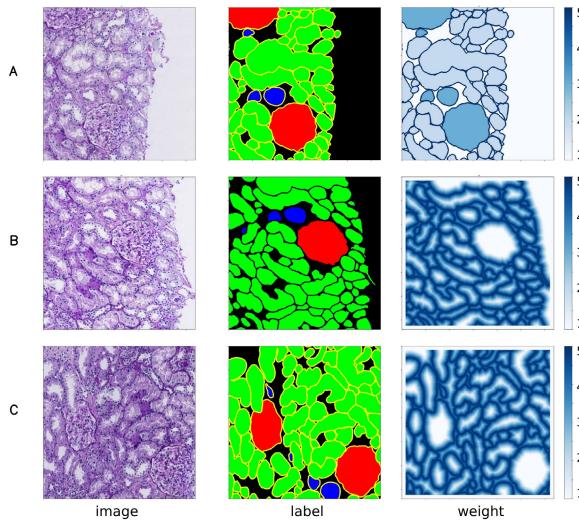


FIGURE 3. The processing method of the target boundary. Three graphs in a row are the original image, label image, weighting factor heat map: A) The target boundary is regarded as a separate class, together with fixed weights. B) The target boundary is not regarded as a separate class, but boundary weights are added. C) The target boundary is regarded as a separate class, and weights are added to the boundary.

this framework, we further developed the Palgo algorithm platform (<https://www.palgo.com.cn/>), which integrates our segmentation models and provides an accessible environment for model deployment and testing. This platform enables users to efficiently apply and validate renal biopsy pathological tissue segmentation algorithms in real-world scenarios. Additionally, we have open-sourced our code and pre-trained models to promote reproducibility and facilitate further research. The code repository is available at: <https://github.com/youngbaby123/Renal-Biopsy-Pathological-Tissue-Segmentation>.

1) Backbone

The backbone had a basic network structure. In recent years, various excellent basic networks (e.g., VGG [50], GoogleNet [51], ResNet [52], DenseNet [53], EfficientNet [48]) and lightweight networks (e.g., MobileNet [54], [55], [56], [57], ShuffleNet [58], [59]) have been proposed. In the experiments, comparisons were made between several latest models and the network models commonly used in medical image analysis, including Unet [42], U²net [60], EfficientNet (b3, b4), Mobilenet-v3, SwinUnet [61] based on SwinTransformer, and ConvWinUnet revised in this study. ConvWinUnet has a similar structure to SwinUnet, except that it replaces the shifted windows in SwinTransformer with group convolution.

2) Neck

The neck [62], [63], [64], [65], [66], [67] connects basic networks with specific tasks and transforms the output of the backbone into the input of the corresponding task by target segmentation according to different tasks such as classification and segmentation. Meanwhile, the neck

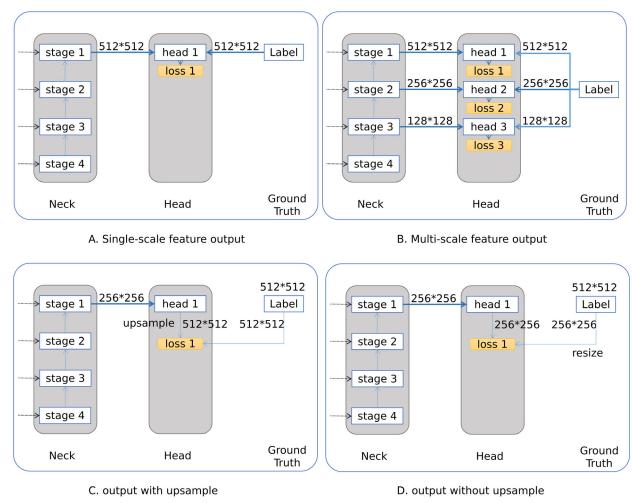


FIGURE 4. The different head structures. A) The structures of single-scale feature output. B) The structures of multi-scale feature output. C) The structures of output with upsampling. D) The structures of output without upsampling.

can fuse the backbone features to make it more suitable for subsequent tasks. In the experiments, the BiFPN and U²net were compared. This study constructed the decoder of SwinUnet with a structure similar to that of the neck and adjusted the position of normalization in the patch, expanding the structure to further improve the model (but this structure is not very competitive with others).

3) Head

The head [68] was set according to the needs of different tasks. In this study, comparative experiments were conducted on single-scale and multi-scale feature outputs. In addition, by using upsampling, this study also tested the impact of the output feature size on the performance of the model. As shown in Fig. 4, in the single-scale feature output, a single head was used to connect the feature output by the neck of the last stage, whereas in the multi-scale feature output, multiple heads were used to connect the feature output by the neck of the corresponding stage. It should be noted that only multi-scale feature outputs were used in the training process. In the test process, only a single specific head was selected as the final output of the model, which could simplify the model and accelerate reasoning. Subsequently, relevant experiments were conducted to select a specific head as the test output. Figure 4C and D show how the size of the output feature map can be changed in combination with upsampling.

C. TRAINING MODULE

According to the description above, the segmentation target belongs to the following five classes: KG1, KT1, KV1, boundary, and background. A unified learning rate and

optimizer were adopted in this study, and Adam was chosen as the optimizer with an initial learning rate of 0.001 and weight decay of 0.0001. A warmup was used for the first 500 iterations, and a step method was used to decay the learning rate. A total of 200 epochs were executed, and the learning rate decayed after 150 epochs. The following two comparative experiments were conducted.

1) Batch normalization and batch size

EfficientNet and U²net were used as backbones for BN [69] and SyncBN [70] to investigate the impact of batch size on the model performance.

2) Loss function

In the experiments, the following losses were compared: CE loss [71], BCE loss [72], Dice loss [73], focal loss [74], SSIM [75], [76], and multi-CE loss. In particular, to verify whether the multi-scale feature output was effective in kidney pathological tissue segmentation, the multi-CE was modified according to the multi-scale feature output combined with CE loss (as shown in Figure 4B).

D. TESTING MODULE

A complete WSI is too large to be simultaneously detected at one time. Thus, during the testing process, the image was cropped into multiple patches for testing and the test results were then spliced to the original size. The complete process is as follows (the sample is shown in Figure 5):

1) Step 1. Crop the image into patches

The circle of the image is filled with a pixel value of 0 and a pixel width of 64 to prevent cross-border crossing. The image was then cropped into patches of size 1024 × 1024 pixels and an overlap of 128 pixels.

2) Step 2. Testing the patch images

All the patches in Step 1 are tested with a trained model.

3) Step 3. Splice the results of the patch images

The middle 896 × 896 pixels of the test result were taken as an effective area and backfilled into the corresponding original image to splice the result. In this process, quantization is performed to optimize memory and speed up the splicing operation.

4) Step 4. Tissue extraction from the spliced results

For the same pixel, the maximum pixel value of the glomerulus, tubules, vessels, boundary, and background scores was taken as the final value of the pixel. The class corresponding to the maximum value was also considered the class of the pixel. Subsequently, the connected domain is extracted from the pixels belonging to the same class as the final target area, and score filtering is performed to eliminate invalid targets.

5) Step 5. Correct the edges of the target

Because the splicing operation causes the boundary target to be incomplete, for the target with various straight lines on the edge, a secondary detection is performed to correct the target with inaccurate edges.

E. PERFORMANCE METRICS

Precision, recall, and F1 scores were used to quantify the performance of the model. Meanwhile, the precision recall curve and receiver operating characteristic curve were used to visualize the actual effects of the different models. It should be noted that because conventional pixel-based statistics (such as Dice) do not consider the integrity of the target organization, this study adopts *mask_IOU* for statistics. We denote the prediction as *mask_p* and the ground truth as *mask_g*. The related performance indicators were calculated using the following formulae:

$$TP = \sum_{i=0}^N \sum_{j=0}^M \phi(\text{mask_IOU}_{ij}) \quad (1)$$

$$FP = N - TP \quad (2)$$

$$FN = M - TP \quad (3)$$

with

$$\begin{aligned} \text{mask_IOU}_{ij} &= \frac{\text{mask_p}_i \cap \text{mask_g}_i}{\text{mask_p}_i \cup \text{mask_g}_i} \\ \phi(a) &= \begin{cases} 0, a < \text{Threshold} \\ 1, a \geq \text{Threshold} \end{cases} \end{aligned}$$

where *Threshold* indicates the IOU threshold, which is set to the default value of 0.7 for all experiments, *N* indicates the number of detected targets, and *M* indicates the number of labeled targets.

IV. RESULTS AND DISCUSSION

This section presents a detailed analysis of the experimental results, which include a comprehensive comparison with previous studies and an in-depth evaluation of the proposed approaches.

Table 1 shows a comparison between this study and previous research. The final improvements based on Unet, U²net, and Transformer architectures demonstrate significant enhancements in recall and precision compared to earlier methods, particularly when compared with instance segmentation-based approaches. As shown in the subsequent analysis, these improvements are closely related to the comprehensive multi-aspect design employed in this study. Figure 6 presents the confusion matrices of the three proposed solutions in this study. It is important to note that these methods are based on semantic segmentation, where non-target regions are classified as background. The confusion matrices reveal that misclassification between categories is minimal, with most errors arising from target regions being classified as background (causing missed detections) or background regions being classified as a target category (causing false positives). Further analysis suggests that these errors may be attributed to annotation inaccuracies, indicating that the actual performance might be better than reflected by the numerical results.

This section presents a detailed analysis of the experimental results, which include four main parts. 1) The results of

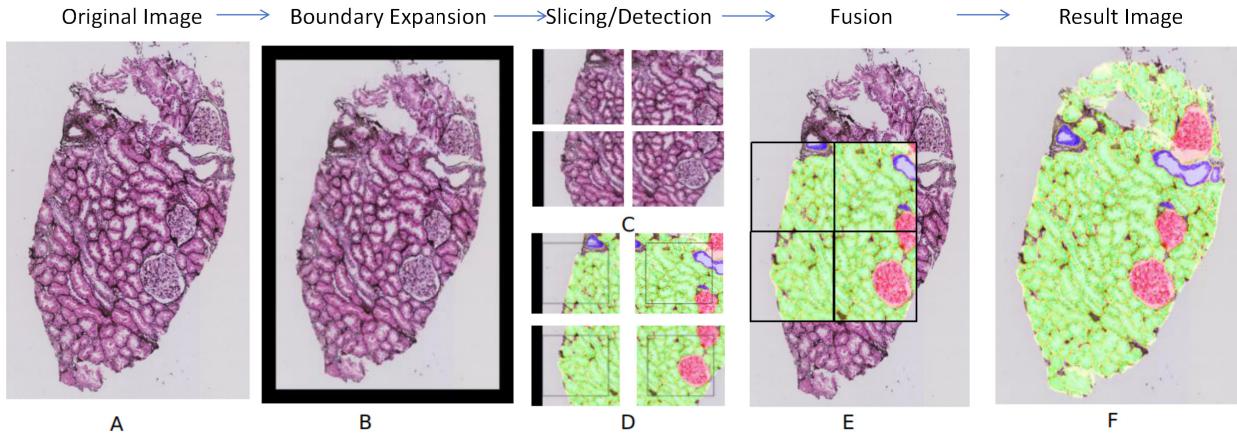


FIGURE 5. Specific flow of the testing process: A) Original image. B) Padding with a pixel value of 0 around the original image. C) Full coverage cut of image B. D) The analysis of the segmentation results through the model and the extraction of the middle effective area. E) The splice of the segmentation results. F) The correction and fusion of the splicing results and the final output results.

TABLE 1. Comparison of performance metrics between proposed methods and existing studies.

Methods	Speed(s)	GFLOPs	Params(M)	Mean F1	KG1		KT1		KV1	
					Precision	Recall	Precision	Recall	Precision	Recall
YOLOv8	310.00	79.00	26.00	53.35	77.58	68.94	77.21	54.33	35.20	17.38
Mask R-CNN	460.00	N/A	N/A	60.46	84.58	72.23	88.73	64.43	45.21	21.13
DeepLabv3	370.00	N/A	N/A	78.73	80.77	92.56	90.72	92.57	50.31	69.28
Unet	380.00	89.00	2.69	80.78	81.51	92.73	91.87	93.54	56.31	71.22
SwinUnet	610.00	97.00	27.00	81.75	81.05	93.67	92.32	93.35	60.79	71.05
ConvWinUnet	410.00	14.00	14.00	82.03	86.53	91.77	92.57	93.31	57.76	71.94
ours(base on Transformer)	630.00	67.00	34.00	84.02	86.30	95.45	92.42	94.04	66.73	69.69
ours(base on Unet)	420.00	26.00	18.00	84.26	86.30	94.97	92.55	94.15	63.88	75.03
ours(base on U ² net)	920.00	150.00	44.00	85.27	86.30	95.45	92.67	94.29	66.40	77.91

the model structure are compared, including the comparison of various types of backbones, necks, and heads, and the analysis of experiments related to BN and batch size. 2) The loss functions for image segmentation were compared, including CE Loss, BCE Loss, Focal Loss, Dice Loss, and SSIM. Simultaneously, multi-CE loss was analyzed and combined with multi-scale feature output. 3) The results of the pre-treatment and post-treatment experiments were compared, including the boundary treatment and weight setting. The input and output feature map sizes were also compared. In addition, a comparative analysis was performed on the finetuning process using the full datasets. 4) The performances of different models were compared in terms of computation speed, calculation amount, and number of parameters.

A. EXPERIMENTAL RESULTS OF THE DIFFERENT BACKBONES

In this study, advanced deep learning models for medical image analysis, including Unet, EfficientNet+Unet, MobileNet+Unet, and U²net were compared. The recently popular transformer structure [77], [78] has also been investigated (i.e., SwinUnet and ConvWinUnet combined with convolutional construction). Based on these results, the ROC curve (Figure 7) and the PR curve (Figure 8)

were plotted, and the optimal F1 values for each class are summarized (shown in Table 2).

It can be seen that the original Unet achieved good results, but there were more false positives, and its boundaries were more blurred. This indicates that Unet, while efficient in capturing general features, struggles with fine boundary details and is prone to over-segmentation errors. The results suggest that the original Unet architecture may need modifications, such as integrating advanced feature extraction techniques, to handle the complexity of renal pathology images effectively.

The Transformer structure also performed well in the comparative study. Its ability to model global contextual information through attention mechanisms contributed to better segmentation results, especially for complex structures like renal vessels. However, the computational cost and slower inference speed of Transformer-based models, as noted in other studies [41], highlight their trade-off between accuracy and efficiency. This aligns with our findings, where despite their robust performance, Transformers require significant resources, making them less practical for real-time or resource-constrained applications.

Overall, the models of U²net and EfficientNet+Unet were better than the other models. This can be attributed to the lightweight yet powerful architecture of U²net, which excels

		A. Base On Transformer				B. Base On UNet				C. Base On U2Net			
		KG1	KT1	KV1	BG	KG1	KT1	KV1	BG	KG1	KT1	KV1	BG
Ground Truth	KG1	95.45	0.21	0.36	3.98	94.97	0.28	0.21	4.53	95.45	0.2	0.32	4.04
	KT1	0.02	94.04	0.1	5.85	0.02	94.15	0.09	5.74	0.02	94.29	0.09	5.6
	KV1	0.13	0.72	69.69	29.47	0.09	0.72	75.03	24.16	0.08	0.63	77.91	21.38
	BG	4.18	54.25	41.57	0	3.87	49.11	47.03	0	4.04	50.46	45.5	0

FIGURE 6. Confusion matrices of the proposed semantic segmentation methods. A) Confusion matrix for the method based on Transformer. B) Confusion matrix for the method based on Unet. C) Confusion matrix for the method based on U²net.

in detecting detailed boundaries, and the superior feature extraction capabilities of EfficientNet when combined with Unet's multi-scale structure. These models not only achieved higher recall and precision but also maintained a good balance between computational efficiency and accuracy, making them particularly suitable for clinical applications.

In addition, Table 2 reveals that there is little difference in the results of renal tubules output by different models, but the results for vessels vary significantly. This is likely due to more complex vessel characteristics. However, further analysis showed that most cases of false-positive (FP) vessels were caused by label deletion. The actual detection effect is much better than that of the exhibited algorithm, but the comparison of the algorithms is fair. Overall, the U²net and EfficientNet+Unet models were the most effective among all the models.

B. EXPERIMENTAL RESULTS OF THE DIFFERENT NECKS & HEADS

The neck is an effective operation for integrating the characteristics of each stage of the network, which can significantly improve the performance of the network. Skip connection structures, such as Unet are common neck operations used in medical image analysis. The head structure is a characteristic extractor with a structured output for handling different tasks. In this study, the use of the decoder of Unet and BiFPN as the neck of EfficientNet is compared. Then, U²net and SwinUnet are compared because the two networks used a neck structure different from Unet. In U²net, the upsampling block uses a U-shaped structure as the base module. SwinUnet creatively uses a transformer structure as the base module and adopts patch expansion as the upsampling layer. Finally, for comparative experiments of the heads, the single-scale feature output and multi-scale feature output are compared.

As presented in Table 3 and Figure 9, it can be seen that 1) BiFPN has no obvious advantages over the Unet decoder. 2) In the model with EfficientNet as the backbone, using upsampling to increase the size of the output feature map can improve the effectiveness of the model, particularly for

vessels with small targets. 3) The multi-scale feature output failed to improve the model effect in EfficientNet, but the gain was obvious in U²net. Moreover, using only the output feature from stage 2 of the neck (multi-scale b2) in the test can achieve the same effect as using the output feature from stage 1 of the neck (multi-scale b1), and can save considerable computational costs. 4) The Transformer-based SwinUnet demonstrates good performance. In this study, the structure of SwinUnet was modified (Figure 9D), and it was found that placing the normalization layer before the linear layer could significantly improve the performance. This indicates that there is still great potential for improving this new structure.

C. EXPERIMENTAL RESULTS OF THE DIFFERENT BN & BATCH SIZE

In this experiment, EfficientNet+Unet and U²net were used as the basic networks to perform ablation experiments on different BN and batch sizes. In the study on EfficientNet+Unet, using SyncBN or not, using multiple batches, and amplifying the output channel size are compared. In the study on U²net, using SyncBN or not and using multiple batches were also compared. In addition, the effect of multi-scale feature output on model performance is also compared.

As shown in Table 4, the use of SyncBN, multiple batches, and amplification of the output channel size significantly improves the performance of EfficientNet. This suggests that these adjustments help stabilize training and enhance feature representation, particularly in deeper networks like EfficientNet. The amplification of output channels likely contributes to better capturing fine details in small structures, such as vessels, which are critical in medical image segmentation tasks.

As shown in Table 5, the use of SyncBN and multiple batches were also effective in improving the performance of the U²net. Multi-stage output is effective for U²net, but its contribution to the performance improvement is not as large as that of multiple batches. This may be attributed to the role of batch size in improving gradient estimation during training and reducing variance.

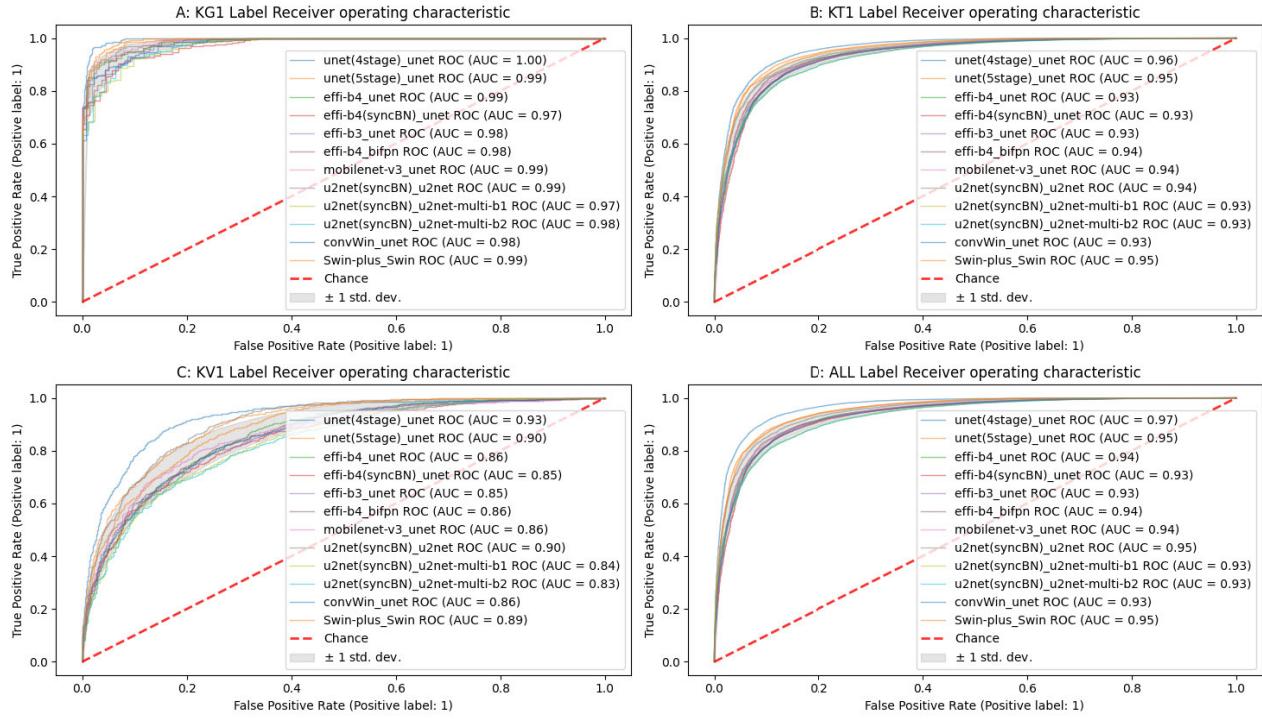


FIGURE 7. The ROC curve of different models. A) is for identifying glomerulus. B) is for identifying tubules. C) is for identifying renal vessels. D) is the summarized ROC curves of different models.

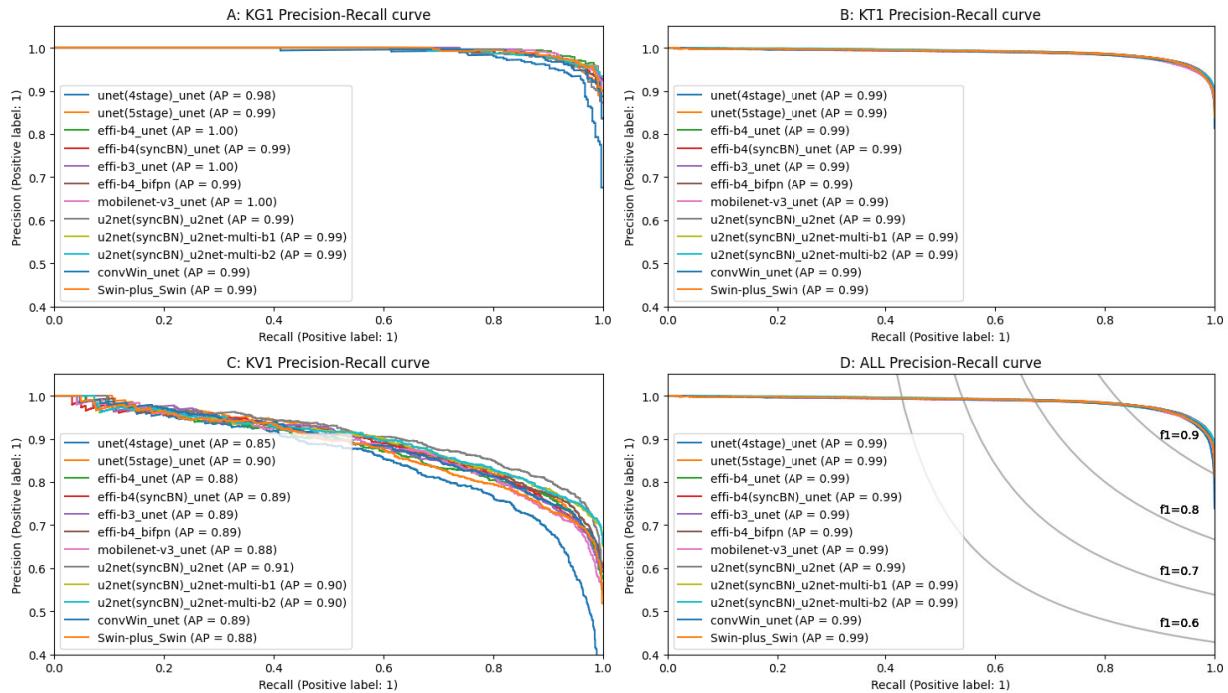


FIGURE 8. The PR curve of different models. A) is for identifying glomerulus. B) is for identifying tubules. C) is for identifying renal vessels. D) is the summarized PR curves of different models.

D. EXPERIMENTAL RESULTS OF THE DIFFERENT LOSS

Several dominant loss function methods were compared in this study, including CE Loss, BCE Loss, Dice Loss,

Focal Loss, and SSIM. As shown in Table 6, BCE and Dice are less effective than the cross-entropy loss. In addition, SSIM slightly improves the model performance,

TABLE 2. The detailed segmentation efficiency of different backbones.

Backbone	Neck&Head	Precision	KG1 Recall	F1	Precision	KT1 Recall	F1	Precision	KV1 Recall	F1
Unet 4stage	Unet	81.51	92.73	86.76	91.87	93.54	92.69	56.31	71.22	62.89
Unet 5stage	Unet	83.33	94.56	88.59	92.84	93.49	93.17	60.98	76.41	67.83
EfficientNet-b4	Unet	86.76	95.48	90.91	91.99	94.58	93.27	56.36	71.23	62.93
EfficientNet-b4 syncBN ^a	Unet	87.90	93.22	90.48	91.97	94.62	93.27	61.68	74.12	67.33
EfficientNet-b3	Unet	86.76	95.48	90.91	91.99	94.58	93.27	56.36	71.23	62.93
EfficientNet-b4	BiFPN	86.99	93.84	90.28	92.71	94.07	93.38	61.96	72.58	66.85
MobileNet-v3	Unet	85.84	94.47	89.95	91.50	93.65	92.56	50.23	71.24	58.92
U ² net syncBN	U ² net	84.93	94.18	89.32	93.35	93.56	93.45	61.59	77.85	68.77
U ² net syncBN	MS b0 ^d	87.21	95.02	90.95	92.88	93.76	93.32	65.42	75.35	70.04
U ² net syncBN	MS b1 ^e	87.21	94.55	90.74	92.89	93.71	93.30	66.96	73.87	70.25
ConvWinUnet ^b	Unet	87.44	93.41	90.33	92.14	93.59	92.86	57.94	73.85	64.94
SwinUnet	SwinUnet	81.05	93.67	86.90	92.32	93.35	92.83	60.79	71.05	65.53
SwinUnet-plusc	SwinUnet	86.30	95.45	90.65	92.42	94.04	93.23	66.73	69.69	68.18

^a syncBN: Replace BN with syncBN in the model.^b ConvWinUnet: The overall structure of ConvWinUnet is similar to that of SwinUnet, except that it replaces the shifted windows in SwinTransformer with group convolution.^c SwinUnet-plus: Based on SwinUnet, the normalization operation of the patch expanding structure is performed before the transformer, and the pre-training model of ImageNet is used for parameter initialization.^d MS b0 represents multi-scale b0: The multi-scale feature output is used in the training process, and the neck output of the last stage is used as the only head input in the testing process.^e MS b1 represents multi-scale b1: The multi-scale feature output is used in the training process, and the neck output of the last second layer is used as the only head input in the testing process.**TABLE 3.** The detailed segmentation efficiency of different necks and heads.

Backbone	Neck&Head	KG1 F1	KT1 F1	KV1 F1	Mean F1
EfficientNet-b4	Unet	90.48	93.27	67.33	83.69
EfficientNet-b4	BiFPN	90.28	93.38	66.85	83.50
EfficientNet-b4	Unet + upsampling ^a	90.43	93.34	69.01	84.26
EfficientNet-b4	Unet + upsampling multi-scale b1 ^b	89.64	93.04	66.77	83.15
U ² net	U ² net	89.32	93.45	68.77	83.85
U ² net	multi-scale b1	90.95	93.32	70.04	84.77
U ² net	U ² net multi-scale b2	90.74	93.30	70.25	84.76
SwinUnet	SwinUnet Normal last ^c	89.29	93.16	67.34	83.26
SwinUnet	SwinUnet Normal first ^d	90.65	93.23	68.18	84.02

^a Unet + upsample: Use upsample in stage 1 of the Unet decoder to double the size of the feature map.^b The multi-scale feature output is used in the training process, and the neck output of stage 1 is used as the only head input in the testing process.^c In the decoder of SwinUnet, the normalization operation of the patch expanding structure is performed after the transformer.^d In the decoder of SwinUnet, the normalization operation of the patch expanding structure was performed before the transformer.**TABLE 4.** The efficiency of using different BN, Batch Sizes, and Output channels in EfficientNet.

Use syncBN	Multi batch	Output channel	KG1 F1	KT1 F1	KV1 F1	Mean F1
✗	✗	✗	56.43	80.07	23.75	53.42
✗	✓	✓	90.65	93.27	63.19	82.37
✓	✗	✗	89.82	92.53	57.40	79.92
✓	✗	✓	90.05	93.24	65.36	82.88
✓	✓	✓	90.48	93.27	67.33	83.69

and there is no statistical improvement when Focal Loss is used.

TABLE 5. The efficiency of using BN, Batch sizes, and multi-stage output in U²net.

Use syncBN	Multi batch	Multi-stage Output	KG1 F1	KT1 F1	KV1 F1	Mean F1
✗	✗	✗	80.88	88.81	42.82	70.84
✗	✓	✗	89.89	93.54	71.20	84.88
✓	✗	✗	89.32	93.45	68.77	83.85
✓	✓	✗	89.45	93.64	71.97	85.02
✓	✗	✓	90.95	93.32	70.04	84.77
✓	✓	✓	90.05	93.60	70.81	84.82

Subsequently, four different staining methods were selected to illustrate the actual segmentation effect of each

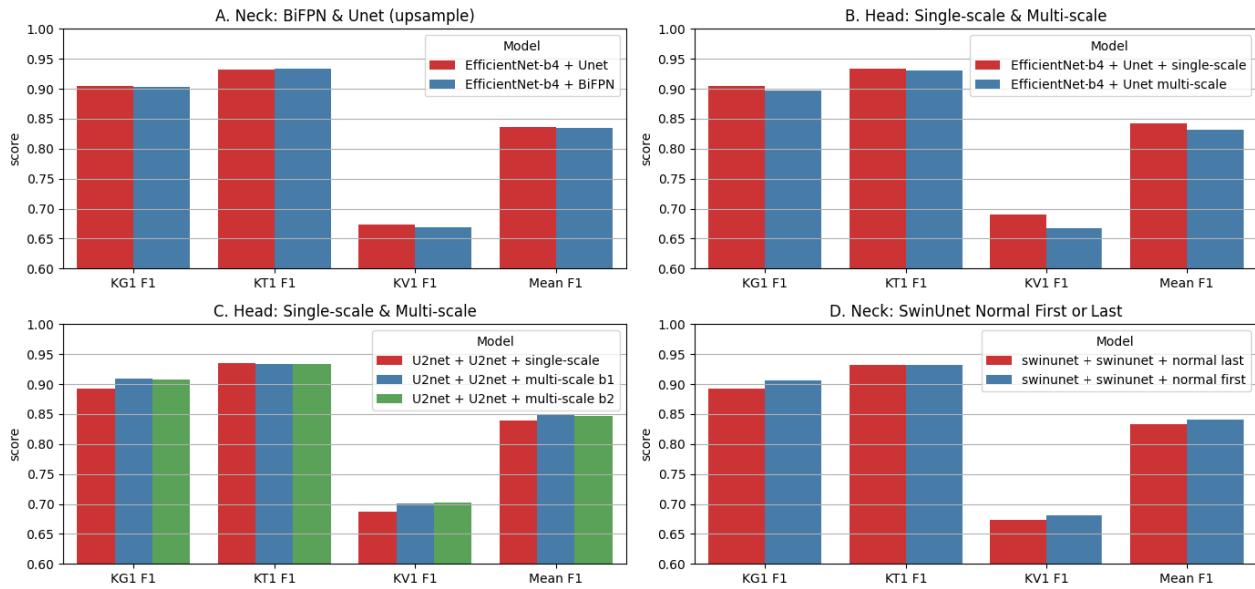


FIGURE 9. The comparison of the neck & head. A) Red, EfficientNet with Unet as the decoder; Blue, EfficientNet with BiFPN as the decoder; B) Red, EfficientNet without multi-stage output; Blue, EfficientNet with multi-stage output; C) Red, U²nets' decoder without multi-stage output; Blue, U²nets' decoder with multi-stage output and the feature of stage 1 is used as the segmentation output; Green, EfficientNet with multi-stage output and the feature of stage 2 is used as the segmentation output; D) Red, placing the linear layer before the normalization layer; Blue, placing the normalization layer before the linear layer.

TABLE 6. The efficiency of different loss functions on EfficientNet.

Loss	KG1 F1	KT1 F1	KV1 F1	Mean F1
CE	89.24	93.42	66.65	83.10
BCE	88.78	92.38	56.07	79.08
CE + Focal	90.37	93.30	65.48	83.05
CE + SSIM	90.24	93.31	66.62	83.39
CE + Dice	89.36	92.42	62.19	81.32

loss function (see Figure 10). It can be seen that the BCE loss and dice loss processed the contours too sloppily, and the effect was obviously weaker than that of other loss functions. In contrast, the SSIM process produces finer contours. This is likely due to the fact that SSIM, unlike other loss functions, emphasizes the structural similarity between the predicted and ground truth images. By focusing on local luminance, contrast, and structure, SSIM encourages the model to better capture spatial dependencies and edge details, which are particularly important in medical image segmentation tasks involving complex and irregular boundaries.

The overall performance improvement of the Focal Loss is not obvious (although it shows a certain effect on segmenting difficult samples). In addition, it is worth noting that the model can detect many targets that are not actually labeled manually, and the segmentation is correct. PAS, PASM, H&E, and Masson staining methods all achieved good results, indicating that the generalization ability of the model was very strong.

TABLE 7. Scale optimization for EfficientNet-b4+Unet model.

Input scale	Output scale	Output channel	KG1 F1	KT1 F1	KV1 F1	Mean F1
0.5	0.5	64	89.82	92.53	57.40	79.92
0.75	0.5	64	90.05	93.24	65.36	82.88
0.75	1	64	89.24	93.42	66.65	83.10
0.75	0.5	32	89.34	93.38	65.40	82.71

TABLE 8. The efficiency of label weights, Boundary label & More data on EfficientNet+Unet model.

Label weights	Boundary label	More data	KG1 F1	KT1 F1	KV1 F1	Mean F1
✗	✗	✗	90.43	93.34	69.01	84.26
✗	✗	✓	94.63	93.06	74.64	87.44
✗	✓	✓	93.75	93.82	77.06	88.21
✓	✓	✓	93.24	93.74	77.66	88.21

E. EXPERIMENTAL RESULTS OF DIFFERENT PRE-TREATMENTS & POST-TREATMENTS

Here, different pre-treatments and post-treatments are discussed, including scales, image pretreatment, and increased training data.

1) Scales

Various experiments were designed to determine the input scale, output scale, and output channel size. Experiments were conducted on different backbones. As shown in Table 7, increasing the input scale, output scale, and output channel size significantly improved

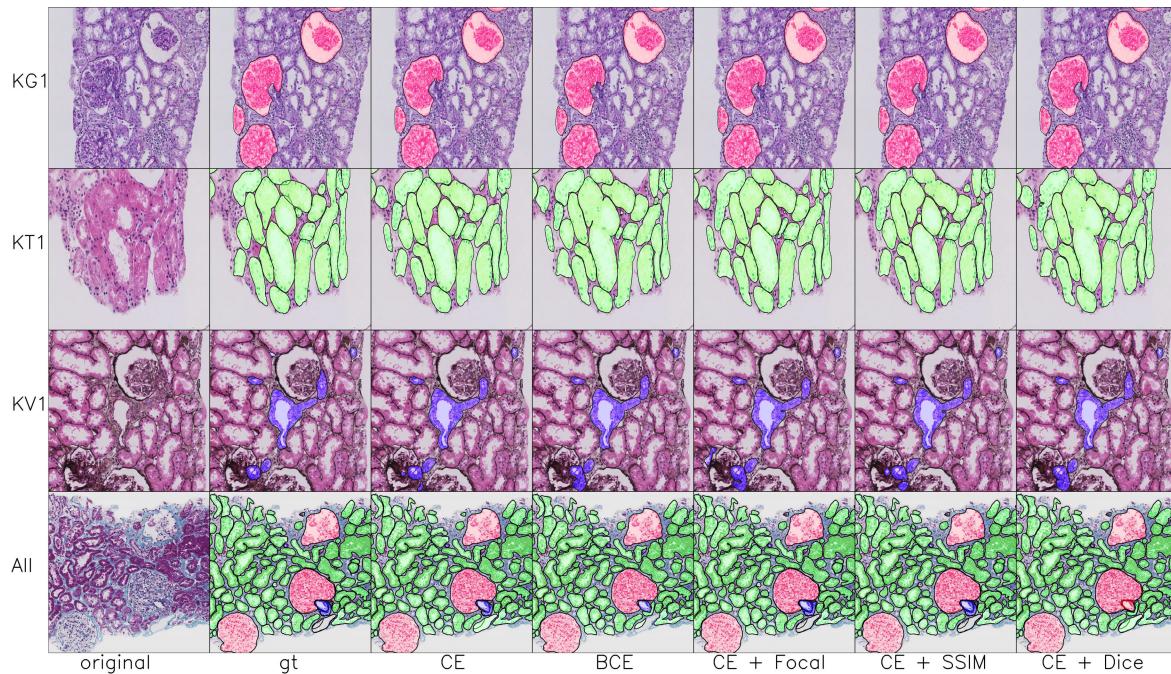


FIGURE 10. The detailed comparison of the segmentation results under different loss functions with different staining methods. The first line shows the segmentation of glomerulus under the use of the PAS staining method (labeled in red); the second line shows the segmentation of tubules under the use of the H&E staining method (labeled in green); the third line shows the segmentation of renal vessels under the use of the PASM staining method (labeled in blue); the fourth line shows the segmentation of all targets under the use of the Masson staining method. Additionally, from left to right, each column presents the original image, the ground truth image, and results under the use of the CE Loss, BCE Loss, CE+Focal Loss, CE+SSIM, and CE+Dice Loss, respectively.

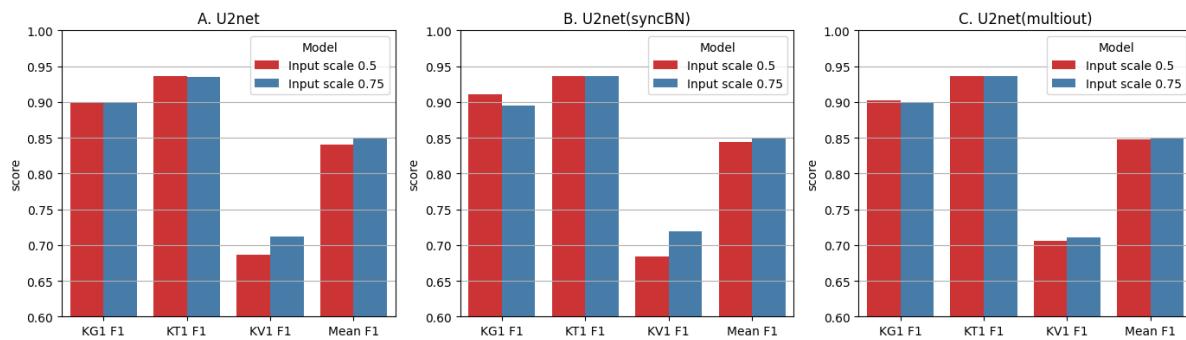


FIGURE 11. The results of comparative experiments on the input scale of U^2 net variants. A) The results of the original U^2 net, where the red bar indicates setting the scale to 0.5, and blue bar indicates setting the scale to 0.75; B) the results of U^2 net+SyncBN, where red bar indicates setting the scale to 0.5, and the blue bar indicates setting the scale to 0.75; C) the results of U^2 net+SyncBN+multi-stage output, where red bar indicates setting the scale to 0.5, and the blue bar indicates setting the scale to 0.75.

the performance of EfficientNet-b4+Unet. Meanwhile, Figure 11 illustrates that increasing the input scale is also effective for U^2 net.

2) Image pretreatment

For image pretreatment, using the boundary as the learning objective and setting the weight of the learning objective both affect the model performance, especially in processing the details of the boundary. The previous content introduced label processing for the boundary and the corresponding boundary weights. It should be noted

that gradient boundary weights were applied to all the models in this study. The label weights here indicate that the gradient boundary weights are multiplied by different weight scales. Herein, the scales are set to 1.5, 1, 2, and 2 for KG1, KT1, KV1, and the boundary. As shown in Table 8, setting the boundary as a separate label significantly improves the model performance, particularly for the segmentation of vessels. Meanwhile, the label weights do not show a significant effect. Artificial interventions did not play a significant role.

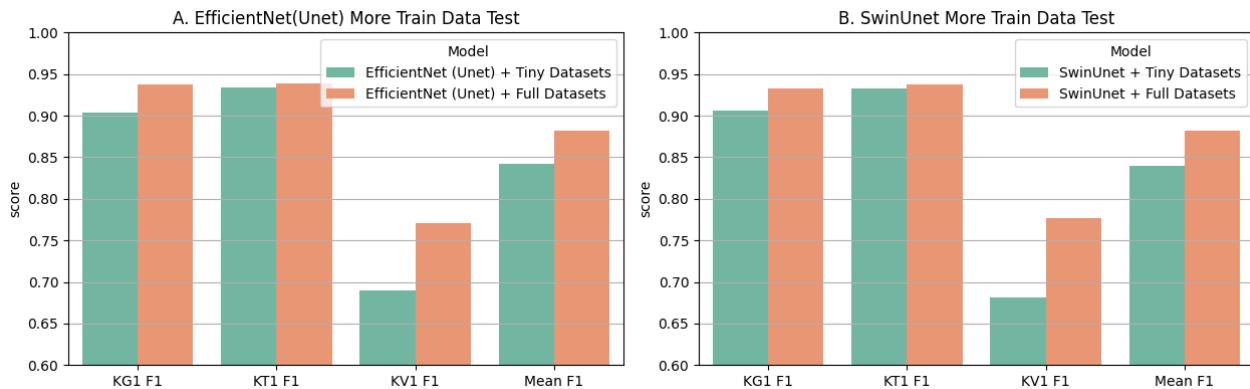


FIGURE 12. The performance comparison under more training samples. A) Using more training samples for EfficientNet+Unet, where the green bar indicates the original training results on the tiny datasets, and the orange bar indicates the training results on the full datasets. B) Using more training samples for SwinUnet, whether the green bar indicates the training results on the tiny datasets, and the orange bar indicates the results on the full datasets.

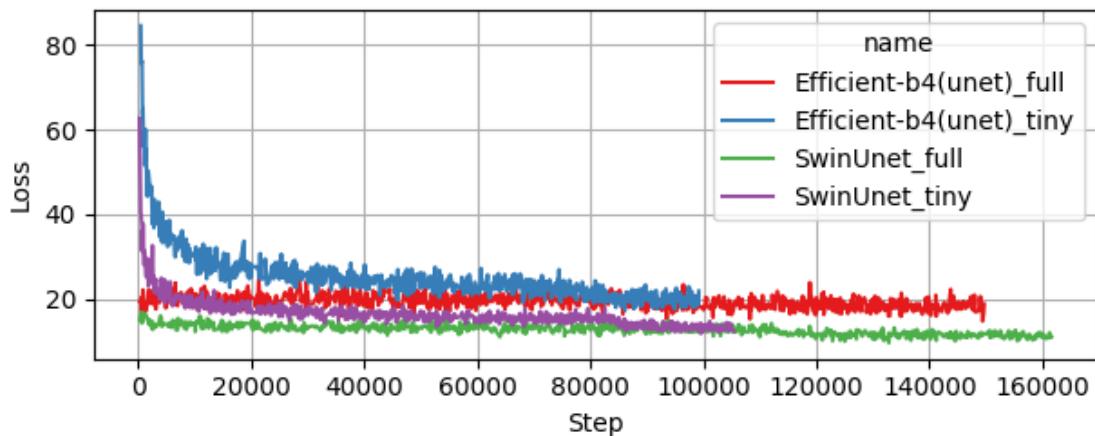


FIGURE 13. The loss function curve in the model training process. The blue and red lines indicate the training curves of EfficientNet-b4+Unet on the tiny datasets and full datasets, respectively; the purple and green lines indicate the training curves of SwinUnet on the tiny datasets and full datasets, respectively.

TABLE 9. The efficiency of label weights, Boundary Label & More data on EfficientNet+Unet model.

Model	Input Scale	Output Scale	Output channel	Mean F1	Speed(s)	GFLOPs	Params(M)
EfficientNet Unet	0.5	0.5	64	79.92	165.96	25.03	17.74
EfficientNet Unet	0.75	0.5	64	82.88	310.78	25.03	17.74
EfficientNet Unet	0.75	1	64	84.26	377.53	25.81	17.74
EfficientNet Unet	0.75	0.5	32	82.71	286.93	18.82	17.65
U ² net	0.75	1	32	83.85	887.23	150.75	44.00
U ² net	0.5	1	32	83.75	462.26	259.69	44.00
U ² net multi-scale b1	0.5	1	32	84.82	449.38	260.23	44.05
U ² net multi-scale b2	0.5	1	32	84.80	434.02	259.69	44.00
SwinUnet	0.75	1	96	84.02	626.71	66.84	34.28

3) Increased training data

As mentioned in the previous section, tiny datasets are simplified from full datasets to validate the effectiveness of the algorithm efficiently and perform more comparative experiments. Here, we compare the models based

on the two datasets. In this study, the training was started from tiny datasets with a total of 200 epochs. Then, the model trained on the tiny datasets was used as a pre-training model, and it was further trained for 30 epochs on the full datasets. As shown in Figure 12, a substantial

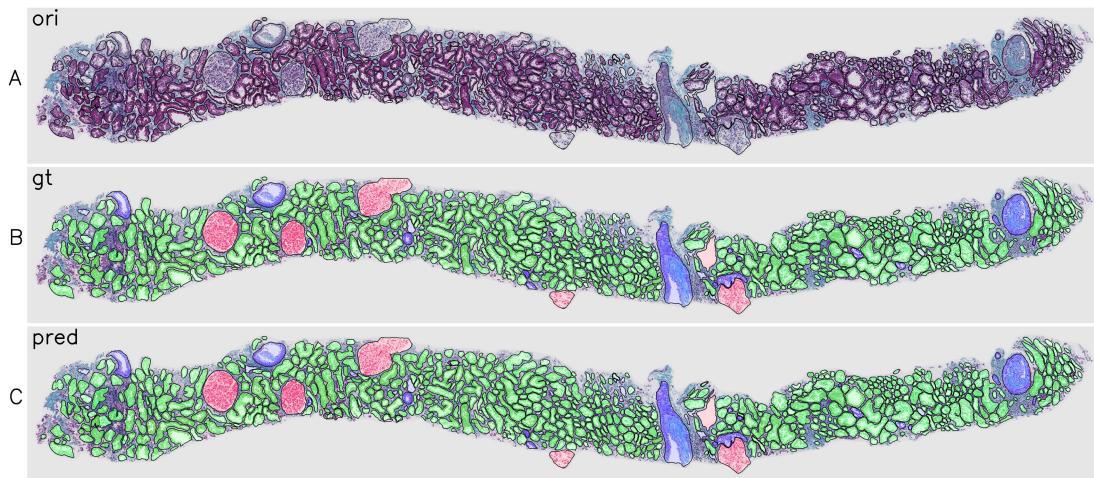


FIGURE 14. The segmentation results of renal pathological tissue: red indicates glomeruli; green indicates tubules; purple indicates renal vessels, and light blue indicates interstitium. Additionally, A) presents the original image; B) presents the ground truth, and C) presents the segmentation results of EfficientNet-b4+Unet.

improvement in model performance was achieved by using full datasets with more training samples. As shown in Figure 13, the 200 epochs on the tiny datasets iterate approximately 100K times, while the 30 epochs on the full datasets iterate approximately 150K times. In addition, the training on the full datasets converges fast and well, suggesting that the model trained on tiny datasets is reliable and easily applied to larger datasets.

F. SPEED, PARAMS, AND GFLOPS

The performance, efficiency, and computing resource consumption of these models for identifying renal pathological tissues were compared. Deep learning methods can identify renal pathological tissues much faster than humans, and faster and more sensitive segmentations enable earlier diagnosis and treatment of patients. In addition, because computing resources are limited, effective segmentation with limited resource consumption is preferred. Hence, in the experiment, EfficientNet+Unet (good performance), U²net, and a transformer (SwinUnet, widely used nowadays) were compared, and the total time required to process 48 images was taken as the statistical indicator of speed.

As shown in Table 9, U²net is slow and requires considerable video memory and computational resources despite its good performance. SwinUnet exhibits good performance in identifying renal pathological tissues; however, it is also slow. EfficientNet+Unet exhibited outstanding performance in all aspects.

V. CONCLUSION

In summary, the experimental results indicate a balanced result that can be achieved by the model structure of EfficientNet-b4+Unet with an input image size of 896 × 896 pixels, multiple batches (four graphics cards, each with a batch size of 2), SyncBN, an input scale of 0.75,

an output scale of 1, and an output channel size of 64. For extreme performance, a model with more parameters such as U²net can be considered. Figure 14 presents the segmentation results of renal pathological tissue obtained using the EfficientNet-b4+Unet model. The visual results demonstrate that this model can effectively segment renal pathological tissues. However, further analysis reveals that vascular segmentation performance is slightly inferior to that of glomeruli and renal tubules. This is primarily due to the presence of small capillary structures and a large number of blank areas in the images, which impact the model's ability to learn vascular features effectively. This indicates potential areas for further optimization.

In this study, the segmentation target was divided into the glomeruli, tubules, and renal vessels. The contributions of this study are as follows.

(1) Complete procedures and solutions for kidney pathology testing are provided, including data processing, model training, and final tests. Each part is described in detail below.

(2) Different algorithms are comprehensively compared, including the classic Unet and the latest SwinUnet, the U²net with high performance, EfficientNet, and MobileNet with high efficiency.

(3) Detailed model settings and subtle algorithm tricks were compared, including a detailed comparison of backbones, necks, heads, loss functions, and algorithm tricks such, as BN, batch size, scale, and multi-stage output.

(4) The best solution is provided by considering the advantages and disadvantages of each algorithm in terms of the model performance, efficiency, and resource consumption. The advantages of EfficientNet+Unet were fully analyzed based on numerical and practical results.

(5) The generalization ability of the models is shown by different staining methods and datasets. the models achieve good segmentation results under different staining methods including H&E, PAS, PASM, and Masson. In addition,

the models showed good generalization ability and strong compatibility from tiny datasets to full datasets.

Despite the significant contributions of this study, several limitations exist that warrant further investigation:

(1) **Model complexity:** While the EfficientNet-b4+Unet model balances performance and efficiency, its computational demands may still limit deployment in resource-constrained environments. Future work could explore lightweight alternatives with comparable performance.

(2) **Manual annotations:** The reliance on manually labeled data introduces potential variability and subjectivity. Incorporating semi-supervised or unsupervised learning approaches could reduce annotation requirements and enhance scalability.

(3) **Specific lesion segmentation:** Although the study focuses on segmenting glomeruli, tubules, and vessels, it does not explicitly address the segmentation of specific pathological features, such as crescents or atrophy. Future research could develop specialized models targeting these specific lesions.

(4) **Real-world integration:** The study primarily evaluates models in a controlled experimental setup. Further research is needed to validate these models in real-world clinical workflows, including integration with diagnostic systems and feedback from pathologists.

By addressing these limitations, future research can build on this study to advance the field of renal pathological tissue segmentation and contribute to more precise and efficient diagnostic tools in clinical practice.

REFERENCES

- [1] J. U. Becker, D. Mayerich, M. Padmanabhan, J. Barratt, A. Ernst, P. Boor, P. A. Cicalese, C. Mohan, H. V. Nguyen, and B. Roysam, "Artificial intelligence and machine learning in nephropathology," *Kidney Int.*, vol. 98, no. 1, pp. 65–75, Apr. 2020.
- [2] I. Girolami, L. Pantanowitz, S. Marletta, M. HermSEN, J. van der Laak, E. Munari, L. Furian, F. Vistoli, G. Zaza, M. Cardillo, L. Gesualdo, G. Gambaro, and A. Eccher, "Artificial intelligence applications for pre-implantation kidney biopsy pathology practice: A systematic review," *J. Nephrol.*, vol. 35, no. 7, pp. 1801–1808, Apr. 2022.
- [3] C. Feng et al., "Artificial intelligence-assisted quantification and assessment of whole slide images for pediatric kidney disease diagnosis," *Bioinformatics*, vol. 40, no. 1, p. 740, Jan. 2024.
- [4] F. P. Schena, V. W. Anelli, D. I. Abbrescia, and T. Di Noia, "Prediction of chronic kidney disease and its progression by artificial intelligence algorithms," *J. Nephrol.*, vol. 35, no. 8, pp. 1953–1971, May 2022.
- [5] P. Krisanapan, S. Tangpanithandee, C. Thongprayoon, P. Patharaditima, and W. Cheungpasitporn, "Revolutionizing chronic kidney disease management with machine learning and artificial intelligence," *J. Clin. Med.*, vol. 12, no. 8, p. 3018, Apr. 2023.
- [6] L. Barisoni, K. J. Lafata, S. M. Hewitt, A. Madabhushi, and U. G. J. Balis, "Digital pathology and computational image analysis in nephropathology," *Nature Rev. Nephrol.*, vol. 16, no. 11, pp. 669–685, Aug. 2020.
- [7] R. S. G. Sealoff, L. H. Mariani, M. Kretzler, and O. G. Troyanskaya, "Machine learning, the kidney, and genotype–phenotype analysis," *Kidney Int.*, vol. 97, no. 6, pp. 1141–1149, Jun. 2020.
- [8] J. Saez-Rodriguez, M. M. Rinschen, J. Floege, and R. Kramann, "Big science and big data in nephrology," *Kidney Int.*, vol. 95, no. 6, pp. 1326–1337, Jun. 2019.
- [9] O. Niel and P. Bastard, "Artificial intelligence in nephrology: Core concepts, clinical applications, and perspectives," *Amer. J. Kidney Diseases*, vol. 74, no. 6, pp. 803–810, Dec. 2019.
- [10] E. Uchino, K. Suzuki, N. Sato, R. Kojima, Y. Tamada, S. Hiragi, H. Yokoi, N. Yugami, S. Minamiguchi, H. Haga, M. Yanagita, and Y. Okuno, "Classification of glomerular pathological findings using deep learning and nephrologist-AI collective intelligence approach," *Int. J. Med. Informat.*, vol. 141, Sep. 2020, Art. no. 104231.
- [11] S. Sheehan, S. Mawe, R. E. Cianciolo, R. Korstanje, and J. M. Mahoney, "Detection and classification of novel renal histologic phenotypes using deep neural networks," *Amer. J. Pathol.*, vol. 189, no. 9, pp. 1786–1796, Sep. 2019.
- [12] P. Boor, "Artificial intelligence in nephropathology," *Nature Rev. Nephrol.*, vol. 16, no. 1, pp. 4–6, Jan. 2020.
- [13] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.
- [14] J. N. Kather, A. T. Pearson, N. Halama, D. Jäger, J. Krause, S. H. Loosen, A. Marx, P. Boor, F. Tacke, U. P. Neumann, H. I. Grabsch, T. Yoshikawa, H. Brenner, J. Chang-Claude, M. Hoffmeister, C. Trautwein, and T. Luedde, "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer," *Nature Med.*, vol. 25, no. 7, pp. 1054–1056, Jun. 2019.
- [15] C. Zeng, Y. Nan, F. Xu, Q. Lei, F. Li, T. Chen, S. Liang, X. Hou, B. Lv, D. Liang, W. Luo, C. Lv, X. Li, G. Xie, and Z. Liu, "Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning," *J. Pathol.*, vol. 252, no. 1, pp. 53–64, Sep. 2020.
- [16] L. Jiang, W. Chen, B. Dong, K. Mei, C. Zhu, J. Liu, M. Cai, Y. Yan, G. Wang, L. Zuo, and H. Shi, "A deep learning-based approach for glomeruli instance segmentation from multistained renal biopsy pathologic images," *Amer. J. Pathol.*, vol. 191, no. 8, pp. 1431–1441, Aug. 2021.
- [17] J. N. Marsh, T.-C. Liu, P. C. Wilson, S. J. Swamidass, and J. P. Gaut, "Development and validation of a deep learning model to quantify glomerulosclerosis in kidney biopsy specimens," *JAMA Netw. Open*, vol. 4, no. 1, Jan. 2021, Art. no. e2030939.
- [18] B. Ginley, B. Lutnick, K.-Y. Jen, A. B. Fogo, S. Jain, A. Rosenberg, V. Walavalkar, G. Wilding, J. E. Tomaszewski, R. Yacoub, G. M. Rossi, and P. Sarder, "Computational segmentation and classification of diabetic glomerulosclerosis," *J. Amer. Soc. Nephrol.*, vol. 30, no. 10, pp. 1953–1967, Oct. 2019.
- [19] Y. Zheng, C. A. Cassol, S. Jung, D. Veerapaneni, V. C. Chitalia, K. Y. M. Ren, S. S. Bellur, P. Boor, L. M. Barisoni, S. S. Waikar, M. Betke, and V. B. Kolachalam, "Deep-learning-driven quantification of interstitial fibrosis in digitized kidney biopsies," *Amer. J. Pathol.*, vol. 191, no. 8, pp. 1442–1453, Aug. 2021.
- [20] M. Salvi, A. Mogetta, A. Gambella, L. Molinaro, A. Barreca, M. Papotti, and F. Molinari, "Automated assessment of glomerulosclerosis and tubular atrophy using deep learning," *Computerized Med. Imag. Graph.*, vol. 90, Jun. 2021, Art. no. 101930.
- [21] M. HermSEN, T. de Bel, M. den Boer, E. J. Steenbergen, J. Kers, S. Florquin, J. J. T. H. Roelofs, M. D. Stegall, M. P. Alexander, B. H. Smith, B. Smeets, L. B. Hilbrands, and J. A. W. M. van der Laak, "Deep learning-based histopathologic assessment of kidney tissue," *J. Amer. Soc. Nephrol.*, vol. 30, no. 10, pp. 1968–1979, Oct. 2019.
- [22] N. Bouteldja, B. M. Klinkhammer, R. D. Bülow, P. Droste, S. Otten, S. V. Stillfried, J. Moellmann, S. Sheehan, R. Korstanje, S. Menzel, P. Bankhead, M. Mietsch, C. Drummer, M. Lehrke, R. Kramann, J. Floege, P. Boor, and D. Merhof, "Deep learning-based segmentation and quantification in experimental kidney histopathology," *J. Am. Soc. Nephrol.*, vol. 32, no. 1, pp. 52–68, Nov. 2020.
- [23] C. P. Jayapandian, Y. Chen, A. R. Janowczyk, M. B. Palmer, C. A. Cassol, M. Sekulic, J. B. Hodgin, J. Zee, S. M. Hewitt, J. O'Toole, and P. Toro, "Development and evaluation of deep learning-based segmentation of histologic structures in the kidney cortex with multiple histologic stains," *Kidney Int.*, vol. 99, no. 1, pp. 86–101, Aug. 2020.
- [24] G. Bueno, M. M. Fernandez-Carrobles, L. Gonzalez-Lopez, and O. Deniz, "Glomerulosclerosis identification in whole slide images using semantic segmentation," *Comput. Methods Programs Biomed.*, vol. 184, Feb. 2020, Art. no. 105273.
- [25] J. J. Hogan, M. Mocanu, and J. S. Berns, "The native kidney biopsy: Update and evidence for best practice," *Clin. J. Amer. Soc. Nephrol.*, vol. 11, no. 2, pp. 354–362, Feb. 2016.
- [26] V. Venkatesh and V. Malaiachamy, "Role of special stains as a useful complementary tool in the diagnosis of renal diseases: A case series study," *Tech. Rep.*, 2019.

- [27] A. B. Farris, J. Vizcarra, M. Amgad, L. A. D. Cooper, D. Gutman, and J. Hogan, "Artificial intelligence and algorithmic computational pathology: An introduction with renal allograft examples," *Histopathology*, vol. 78, no. 6, pp. 791–804, May 2021.
- [28] J. Gallego, A. Pedraza, S. Lopez, G. Steiner, L. Gonzalez, A. Laurinavicius, and G. Bueno, "Glomerulus classification and detection based on convolutional neural networks," *J. Imag.*, vol. 4, no. 1, p. 20, Jan. 2018.
- [29] O. Sarkar, M. R. Islam, M. K. Syfullah, M. T. Islam, M. F. Ahamed, M. Ahsan, and J. Haider, "Multi-scale CNN: An explainable AI-integrated unique deep learning framework for lung-affected disease classification," *Technologies*, vol. 11, no. 5, p. 134, Sep. 2023, doi: [10.3390/technologies11050134](https://doi.org/10.3390/technologies11050134).
- [30] M. F. Ahamed, M. M. Hossain, M. Nahiduzzaman, M. R. Islam, M. R. Islam, M. Ahsan, and J. Haider, "A review on brain tumor segmentation based on deep learning methods with federated learning techniques," *Computerized Med. Imag. Graph.*, vol. 110, Dec. 2023, Art. no. 102313, doi: [10.1016/j.compmedimag.2023.102313](https://doi.org/10.1016/j.compmedimag.2023.102313).
- [31] M. F. Ahamed, M. R. Islam, M. Nahiduzzaman, M. J. Karim, M. A. Ayari, and A. Khandakar, "Automated detection of colorectal polyp utilizing deep learning methods with explainable AI," *IEEE Access*, vol. 12, pp. 78074–78100, 2024, doi: [10.1109/ACCESS.2024.3402818](https://doi.org/10.1109/ACCESS.2024.3402818).
- [32] M. Faysal Ahamed, M. Robiul Islam, T. Hossain, K. Syfullah, and O. Sarkar, "Classification and segmentation on multi-regional brain tumors using volumetric images of MRI with customized 3D U-Net framework," in *Proc. Int. Conf. Inf. Commun. Technol. Develop.* Singapore: Springer, 2023, pp. 223–234, doi: [10.1007/978-981-19-7528-8_18](https://doi.org/10.1007/978-981-19-7528-8_18).
- [33] M. F. Ahamed, M. Nahiduzzaman, M. R. Islam, M. Naznine, M. Arselene Ayari, A. Khandakar, and J. Haider, "Detection of various gastrointestinal tract diseases through a deep learning method with ensemble ELM and explainable AI," *Exp. Syst. Appl.*, vol. 256, Dec. 2024, Art. no. 124908, doi: [10.1016/j.eswa.2024.124908](https://doi.org/10.1016/j.eswa.2024.124908).
- [34] M. F. Ahamed, F. B. Shafi, M. Nahiduzzaman, M. A. Ayari, and A. Khandakar, "Interpretable deep learning architecture for gastrointestinal disease detection: A tri-stage approach with PCA and XAI," *Comput. Biol. Med.*, vol. 185, Feb. 2025, Art. no. 109503, doi: [10.1016/j.combiomed.2024.109503](https://doi.org/10.1016/j.combiomed.2024.109503).
- [35] M. F. Ahamed, M. M. Hossain, M. M. Mary, and M. R. Islam, "Optimizing skin lesion segmentation with UNet and attention-guidance utilizing test time augmentation," in *Proc. 6th Int. Conf. Electr. Eng. Inf. Commun. Technol. (ICEEICT)*, May 2024, pp. 568–573, doi: [10.1109/iceeict2016.2024.1053452](https://doi.org/10.1109/iceeict2016.2024.1053452).
- [36] M. F. Ahamed, A. Salam, M. Nahiduzzaman, M. Abdullah-Al-Wadud, and S. M. R. Islam, "Streamlining plant disease diagnosis with convolutional neural networks and edge devices," *Neural Comput. Appl.*, vol. 36, no. 29, pp. 18445–18477, Oct. 2024, doi: [10.1007/s00521-024-10152-y](https://doi.org/10.1007/s00521-024-10152-y).
- [37] M. M. T. Titu, M. M. Mary, M. F. Ahamed, T. M. Oishee, and M. M. Hasan, "Utilizing customized 3D U-Net framework for the classification and segmentation of multi-regional brain tumors in volumetric MRI images," in *Proc. 3rd Int. Conf. Advancement Electr. Electron. Eng. (ICAEEE)*, Apr. 2024, pp. 1–6, doi: [10.1109/icaeee62219.2024.10561700](https://doi.org/10.1109/icaeee62219.2024.10561700).
- [38] Y. Bae, J. Byun, H. Lee, and B. Han, "Comparative analysis of chronic progressive nephropathy (CPN) diagnosis in rat kidneys using an artificial intelligence deep learning model," *Toxicological Res.*, vol. 40, no. 4, pp. 551–559, Oct. 2024.
- [39] G. M. Dimitri, P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, F. Scarselli, A. Zacchi, G. Garosi, T. Marcuzzo, and S. A. Tripodi, "Deep learning approaches for the segmentation of glomeruli in kidney histopathological images," *Mathematics*, vol. 10, no. 11, p. 1934, Jun. 2022.
- [40] S. Hara, E. Haneda, M. Kawakami, K. Morita, R. Nishioka, T. Zoshima, M. Kometani, T. Yoneda, M. Kawano, S. Karashima, and H. Nambo, "Evaluating tubulointerstitial compartments in renal biopsy specimens using a deep learning-based approach for classifying normal and abnormal tubules," *PLoS ONE*, vol. 17, no. 7, Jul. 2022, Art. no. e0271161.
- [41] X. Feng, T. Wang, X. Yang, M. Zhang, W. Guo, and W. Wang, "ConvWin-UNet: UNet-like hierarchical vision transformer combined with convolution for medical image segmentation," *Math. Biosciences Eng.*, vol. 20, no. 1, pp. 128–144, 2022.
- [42] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Nov. 2015, pp. 234–241.
- [43] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.
- [44] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1055–1059.
- [45] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U²-net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107404.
- [46] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*.
- [47] M. Y. Lu, D. F. K. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, "Data-efficient and weakly supervised computational pathology on whole-slide images," *Nature Biomed. Eng.*, vol. 5, no. 6, pp. 555–570, Mar. 2021.
- [48] Y. Xu, Z. Jia, L.-B. Wang, Y. Ai, F. Zhang, M. Lai, and E. I.-C. Chang, "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features," *BMC Bioinf.*, vol. 18, no. 1, p. 281, Dec. 2017.
- [49] K. Chen et al., "MMDetection: Open MMLab detection toolbox and benchmark," 2019, *arXiv:1906.07155*.
- [50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [51] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [52] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [53] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [54] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [55] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
- [56] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.
- [57] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le, "MnasNet: Platform-aware neural architecture search for mobile," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2820–2828.
- [58] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.
- [59] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet v2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jan. 2018, pp. 116–131.
- [60] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*.
- [61] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-UNet: UNet-like pure transformer for medical image segmentation," 2021, *arXiv:2105.05537*.
- [62] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
- [63] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
- [64] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jan. 2018, pp. 1055–1059.

- [65] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," 2015, *arXiv:1511.00561*.
- [66] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.
- [67] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10781–10790.
- [68] X. Dai, Y. Chen, B. Xiao, D. Chen, M. Liu, L. Yuan, and L. Zhang, "Dynamic head: Unifying object detection heads with attentions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7373–7382.
- [69] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.
- [70] C. Peng, T. Xiao, Z. Li, Y. Jiang, X. Zhang, K. Jia, G. Yu, and J. Sun, "MegDet: A large mini-batch object detector," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6181–6189.
- [71] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Operations Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005.
- [72] U. Ruby and V. Yendapalli, "Binary cross entropy with deep learning technique for image classification," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 5393–5397, Aug. 2020.
- [73] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.
- [74] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [75] Z. Wang, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 604–606, Apr. 2004.
- [76] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, 2003, pp. 1398–1402.
- [77] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [78] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.



XINXIN HE received the M.A. degree from Hangzhou Medical College, Hangzhou, China, in 2024.

She is currently an Experimental Technician with the Nephrology Department, Zhejiang Provincial Hospital of Traditional Chinese Medicine. Her research interest includes the therapeutic advancements of traditional Chinese medicine monomers in treating diabetic nephropathy. She has also contributed to the publication of seven SCI articles.



WEI WANG received the bachelor's degree from Shaoxing University of Arts and Sciences, in 2023.

He is currently a Resident Physician specializing in diagnostic ultrasound and interventional treatments. He is primarily responsible for kidney biopsy punctures and challenging procedures, such as thyroid and liver punctures. His research interests include the integration of artificial intelligence with ultrasound technology for applications in nephrology and oncology. He is a member of Zhejiang Shuli Medical Association, Ultrasound Branch.



QIANG HE received the M.D. degree from Zhejiang University School of Medicine, Hangzhou, China, in 2003.

He is currently a Chief Physician, a Professor, a Doctoral Supervisor, and the Secretary of the Party Committee, Zhejiang Provincial Hospital of Traditional Chinese Medicine. He leads the Key Discipline of Renal Immunology and the Key Discipline of Blood Purification in Traditional Chinese Medicine in Zhejiang and he is the Director of the Key Laboratory of Traditional Chinese Medicine Kidney Disease. He has extensive experience in kidney transplantation and nephrology, with over 50 SCI articles published as the first or corresponding author. His research interests include clinical immune tolerance in transplant kidneys, mechanisms of immune molecules in organ transplantation, urinary biomarkers for renal injury, high-quality dialysis, blood purification techniques, minimally invasive techniques, peritoneal dialysis, and new immunosuppressants.



JUAN JIN received the M.D. degree from Zhejiang University School of Medicine, Hangzhou, China, in 2020.

She is currently a Chief Physician, a Professor, a Doctoral Supervisor, and the Deputy Director of Zhejiang Provincial Hospital of Traditional Chinese Medicine. She is dedicated to researching the mechanisms of diabetic nephropathy and the development of renal clear cell carcinoma, with over 50 SCI articles published as the first or corresponding author. Her research interests include the mechanisms and clinical treatment of refractory kidney disease, the exploration of novel immune biomarkers for renal injury, and the epigenetic mechanisms of renal cancer. She is a Youth Member of Chinese Medical Doctor Association, Nephrology Branch, and Zhejiang Medical Association, Nephrology Branch, and an Executive Member of Chinese Geriatric Health Medical Research Association, Geriatric Kidney Disease Prevention and Treatment Branch.



YAN LIU received the master's degree from Shanxi Medical University, China, in 2017.

She is currently a Resident Physician with Zhejiang Provincial Hospital of Traditional Chinese Medicine, specializing in the pathological diagnosis of kidney and tumor diseases and contributing to the development of the hospital's nephrology department. Her research interests include the pathological diagnosis of kidney diseases and tumors, with a particular interest in applying and

exploring artificial intelligence techniques in nephrology and oncology. She is a member of Chinese Medical Equipment Association, Pathology Branch.



TAIPING WANG received the Ph.D. degree from Macau University of Science and Technology, Macau, China, in 2024.

He is currently the Research and Development Head of Hangzhou Medipath Intelligent Technology Company Ltd., and leads the Nephrology Pathology AI Project, particularly focusing on transplant kidney pathology. He is proficient in machine vision technologies and applications, holding nine invention patents as the first inventor.