ORIGINAL ARTICLE

NEPHROLOGY · APSN · WILEY

Deep learning radiomics based on ultrasound images for the assisted diagnosis of chronic kidney disease

Shuyuan Tian [1,2] | Yonghong Yu [2] | Kangjian Shi [3] | Yunwen Jiang [2] | Huachun Song [2] | Yuting Wang [2] | Xiaoqian Yan [4] | Yu Zhong [4] | Guoliang Shao [5]

[1]The Second School of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou, PR China

[2]Department of Ultrasound, Tongde Hospital of Zhejiang Province, Hangzhou, PR China

[3]College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, PR China

[4]Department of Nephropathy, Tongde Hospital of Zhejiang Province, Hangzhou, PR China

[5]Department of Radiology, Zhejiang Cancer Hospital, Hangzhou, PR China

Correspondence
Guoliang Shao, Department of Radiology, Zhejiang Cancer Hospital, Hangzhou 320022, PR China.
Email: shaoguoliang67@163.com

Abstract

Aim: This study aimed to explore the value of ultrasound (US) images in chronic kidney disease (CKD) screening by constructing a CKD screening model based on grey-scale US images.

Methods: According to the CKD diagnostic criteria, 1049 patients from Tongde Hospital of Zhejiang Province were retrospectively enrolled in the study. A total of 4365 renal US images were collected from these patients. Convolutional neural networks were used for feature extractions and a screening model was constructed by fusing ResNet34 and texture features to identify CKD and its stage. A comparative analysis was performed to compare the diagnosis results of the model with physicians.

Results: When diagnosing CKD or non-CKD, the receiver operating characteristic curve (AUC) of our model was 0.918 and that of the senior physician group was 0.869 ($p < .05$). For the diagnosis of CKD stage, the AUC of our model for CKD G1–G3 was 0.781, 0.880, and 0.905, respectively, while the AUC of the senior physician group for CKD G1–G3 was 0.506, 0.586, and 0.796, respectively; all differences were statistically significant ($p < .05$). The diagnostic efficiency of our model for CKD G4 and G5 reached the level of the senior physicians group. Specifically, the AUC of our model for CKD G4-G5 was 0.867 and 0.931, respectively, while the AUC of the senior physician group for CKD G4-G5 was 0.838 and 0.963, respectively (all $p > .05$).

Conclusions: Our deep learning radiomics model is more effective than senior physicians in the diagnosis of early CKD.

KEYWORDS
chronic kidney disease, convolutional neural network, radiomics, ultrasound

Summary at a glance

We constructed a deep learning radiomics model based on grey-scale US images by fusing ResNet34 and texture features to identify CKD and its stage. Our deep learning radiomics model was more effective than senior physicians in the diagnosis of early CKD.

1 | INTRODUCTION

Chronic kidney disease (CKD) is a widespread public health problem affecting more than 12% of the population worldwide.[1] Over the past few decades, the incidence of morbidity in patients with CKD has increased dramatically, posing substantial medical and economic burdens on the global healthcare system.[2–4] CKD often develops with no obvious symptoms and has often progressed to uremia by the time medical care is sought. However, if CKD can be detected and treated early, the disease can be well controlled and even reversed. Studies have shown that the early diagnosis and prevention of CKD can delay the progression of kidney damage, and accurate staging of CKD is one key factor in this.[5]

CKD is defined as abnormalities of the structure or function of the kidneys, present for >3 months, with implications for health. According to the estimated glomerular filtration rate(eGFR) value, CKD can be divided into five stages: G1-G5.[6] At present, the most commonly used CKD screening methods include testing the urine for protein and testing the blood for serum creatinine. However, biochemical tests of blood and urine are tedious and time-consuming. Moreover, people often ignore these tests during health checkups, missing the early detection of CKD.[7] Pathological examination of CKD can determine the degree and cause of kidney fibrosis, but this requires a renal biopsy, which is invasive and can have adverse side effects such as perirenal hematoma, arteriovenous fistula, or infection.[8] Furthermore, it is not feasible to repeat a biopsy on the same patient for dynamic surveillance and clinical management during longitudinal follow-up. In contrast, conventional ultrasound (US) is radiation-free, cost-effective, noninvasive in nature, and is the first-line imaging modality for CKD diagnosis. Most renal diseases are first detected by US.[9] In clinical practice, this method can be used by experienced physicians to easily identify end-stage CKD with renal atrophy or significant changes on renal echo. However, for the vast majority of patients with early-stage CKD or transitional kidney disease, this method is beyond the capacity of human vision and is difficult to perform in clinical practice.[10]

Radiomics is an emerging technology that offers high-throughput extraction of large amounts of imaging features from medical imaging data. This approach has received significant attention recently due to its ability to extract imaging features that are invisible to the naked eye.[11,12] Radiomics employs two artificial intelligence (AI) analytical strategies, that is, machine learning and deep learning, to transform a large amount of minable ROI(region of interest)-based image data into characteristic and representative radiomics features. Compared with traditional US examinations, US radiomics can obtain more information on subtle lesions, greatly improving the accuracy of disease diagnosis.[13] Published studies have demonstrated the promising prospects of radiomics features derived from US images for disease diagnosis,[14] prognostic evaluation,[15] and treatment guidance.[16]

In the present study, a deep learning radiomics (DLR) model based on grey-scale US images of the kidney was applied to the diagnosis and staging of CKD. This model was compared with the diagnostic efficiency of radiologists to explore the application value of grey-scale US imaging in the diagnosis of CKD.

2 | MATERIALS AND METHODS

2.1 | Study population

This was a single-center, retrospective study. In order to avoid possible prejudice in the research process, private information, such as the patient's identity and occupation, was kept confidential when collecting the renal US image data. The algorithm used for modelling in this research is open to all developers and users. The study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee of Tongde Hospital of Zhejiang Province. The requirement for informed consent was waived owing to the retrospective study design. This study followed the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines for diagnostic studies.

From April 2018 to December 2022, 870 patients with CKD and 179 patients with non-CKD were identified at the study hospital and were included in the study.

The inclusion criteria were as follows: ① for CKD patients, diagnosis according to the Kidney Disease Improving Global Outcomes (KDIGO) 2012 clinical practice guidelines[17]; ② complete ultrasonic renal imaging data for the patient could be retrieved from the PACS system of the hospital; ③ the laboratory indices of renal function within 24 h prior to ultrasonic examination of the patient's kidneys could be obtained from the hospital inspection system; ④ all two-dimensional US kidney images were acquired from the LOGIQ-E9 ultrasonic diagnostic apparatus (GE Company). The exclusion criteria were as follows: ① patients with a solitary kidney or unilateral nephrectomy; ② patients with obvious solid or cystic nodules on renal US; ③ unqualified US images below the collection standard for kidney US images; ④ ultrasonic images that met the collection

standard but had poor image quality which may adversely affect the diagnostic ability.

Finally, a total of 4365 renal US images, including 756 normal renal images, 597 renal images in the G1 stage, 696 renal images in the G2 stage, 1217 renal images in the G3 stage, 536 renal images in the G4 stage, and 563 renal images in the G5 stage, were included in the data set for this study. The images were divided into a training set, verification set, and test set at a ratio of 8:1:1. Images from the same patient were assigned to the same data set during image assignment. After training and verification, the constructed deep learning model was used for diagnosis and was compared with the diagnoses of US professionals.

2.2 | Imaging acquisition and region of interest segmentation

A LOGIQ-E9 ultrasonic diagnostic apparatus (GE, USA) equipped with a C1–6 probe (1–6 MHz) was used for renal US examinations. Examinations were performed by one of four board-certified abdominal radiologists. Before each examination, the proper contrast mode, including gain, depth, acoustic window, and focal zone, were set. Each patient lay on the examination bed in the supine position during the examination. Grayscale US images of the kidneys at the coronal section were captured from each participant.

Two board-certified abdominal radiologists and a professional modeller in the research group established a semantic annotation database of kidney US images with a resolution of 1552×970 pixels. To enhance the efficiency of modelling and reduce the interference from irrelevant information, the US images of the renal cortex as ROIs were separated by the semantic segmentation network U-Net. Further, all segmented US images were tested once and arranged from high to low according to the segmentation effect. These were then sequentially fed into the recurrent neural network (RNN) for further augmentation. The image features with lower difficulty were grouped with those with higher difficulty to optimize the high-difficulty samples and generate more accurate semantic segmentations. Lastly, the ROIs were cropped and unified to a size of 224×224 as inputs for the deep learning radiomics model training.

2.3 | Development of the deep learning radiomics model

As shown in Figure 1, the deep learning radiomics model contained two parrel branches of deep feature extraction and texture feature extraction. The deep feature extraction focused on capturing the global information of the images. This was implemented by a standard deep convolutional neural network (CNN). On the other hand, the texture feature extraction captured subtle local spatial details. This was accomplished by the grey level co-occurrence matrix (GLCM) and the histogram of oriented gradient (HOG). The global and local feature information, which are complementary and diverse, were fused to further enhance the expression ability of the model.

ResNet, a classic deep CNN, adopts residual connection structures to facilitate deeper feature representation for network learning. It exhibits powerful and efficient feature extraction capabilities. In this work, ResNet-34 was used to extract the deep features of the ROIs of the kidney, with 34 learnable layers in total. Specifically, a convolutional layer with 64 channels and 7×7 kernel size was first utilized to extract the initial features. This was followed by a 3×3 max pooling layer with stride 2 to eliminate redundant information and further reduce the size of the feature maps to 56×56. Then, a series of residual blocks with skip connections were utilized to enhance the extraction of deep features and gradually reduce the size of the feature maps. Finally, a feature map with a channel dimension of 512 and a size of 7×7 was obtained. This was further reduced to a 512-dimensional vector by average pooling. The detailed configuration of ResNet-34 is shown in Figure 2.

GLCM considers the spatial relationship of the grey levels of pixels in the image. The calculation of GLCM involves distances and angles to reflect the comprehensive information of the image in terms of orientation and spacing.[18,19] Here, a total of 16 GLCMs were calculated at distances 1, 2, 3, and 4 and directions $0°$, $45°$, $90°$, and $135°$, where each GLCM was used for a series of second-order texture calculations. These features were computed based on five commonly used statistical measures, namely, contrast, dissimilarity, homogeneity, energy, and autocorrelation, to form the 80-dimensional texture features of the image to be analysed.

HOG is another widely used texture descriptor that makes use of the image gradient statistics to reflect the edge of a local object. To extract the features, the ROI image was divided into multiple small square regions called cells. In each cell, a histogram of oriented gradients was computed to form the corresponding texture description.[20] These cells together formed a larger connected region block. We concatenated the eigenvectors of all cells and normalized them to obtain the texture feature of the block. Furthermore, the HOG texture feature vector (324 dimensions in this work) of the full image was realized by combining the gradient statistics of all blocks.

GLCM and HOG consider the texture information of images from different perspectives; these approaches are compatible and beneficial for extracting more accurate descriptions and analyses. Therefore, we concatenated the feature vectors of both to obtain a more robust representation of the texture features. Moreover, a texture branch with a residual structure was designed to further enhance the texture features. Specifically, three dense layers and skip connections were involved, where each of the dense layers consisted of a linear layer, batch normalization, ReLU activation function, and dropout strategy, with a dropout rate of 0.5. The number of neurons in each dense layer was kept the same as the dimensionality of the input fused texture features, that is, 404 dimensions. Importantly, the design of the skip connections in the texture branch not only ensured the integrity of the original texture feature but also captured additional information to enrich the diversity.
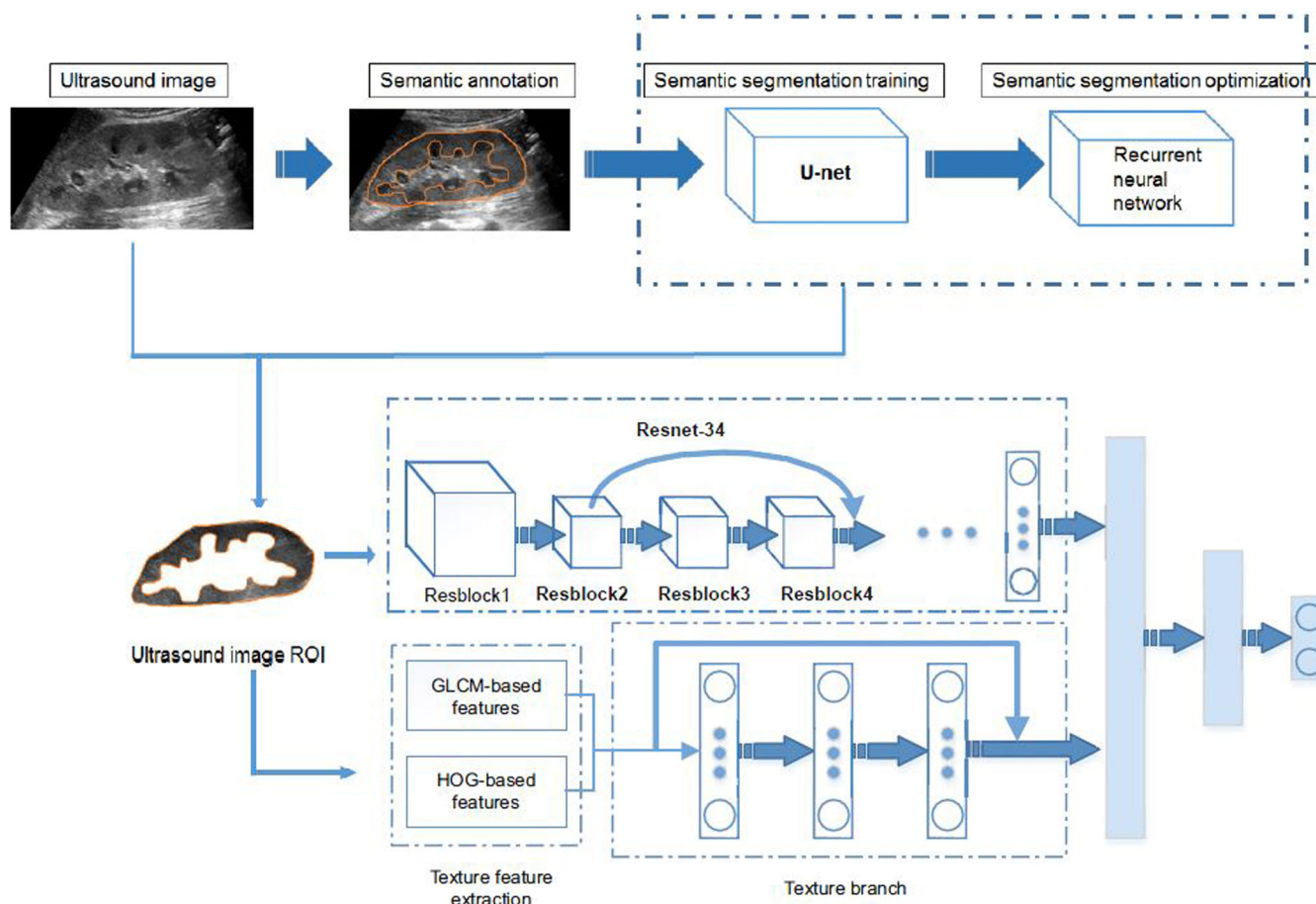
FIGURE 1 Workflow of our model. An overview of the deep learning radiomics model, where ResNet-34 was used as the backbone of the deep convolutional neural network to extract deep (global) features from ROI images, and grey level co-occurrence matrix (GLCM) and histogram of oriented gradient (HOG) considered spatial information to capture diverse texture (local) features.

ResNet-34	Input size	224x224				
	Structure	Conv_1	Conv_2	Conv_3	Conv_4	Conv_5
		$7 \times 7, 64$, stride=2 , 3×3 max pool, stride=2	$\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 3 \times 3, & 256 \\ 3 \times 3, & 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix} \times 3$
	Output size	56×56	56×56	28×28	14×14	7×7

FIGURE 2 Configuration of ResNet-34.

After obtaining the deep (global) features and the texture (local) features, we further fused the two feature vectors to attain a more comprehensive feature representation that was strongly expressive and discriminative. The fused feature representation was passed into a two-layer dense block with 256 hidden neurons for the final classification tasks. Cross entropy loss was utilized as the training objective to optimize the model.

Implementation Details: The stochastic gradient descent (SGD) optimizer with a momentum of 0.9 and weight decay of 2×10^{-4} was adopted to optimize the model training for a total of 40 epochs. For the sake of efficiency, we set the mini-batch size to 64 as the standard training procedure. In addition, the initial learning rate was set to 0.1,

and this was multiplied by 0.2 every eight epochs. All experiments were implemented by Pytorch on GeForce RTX2080Ti GPU with 11 GB memory. The codes are available at https://github.com/cloak-s/DLRM.

2.4 | Diagnosis of professional radiologists

In order to compare the diagnostic performance of the deep learning model with that of different professional doctors, the diagnostic task was completed on the same test data set. The junior professional physician group consisted of three professional, certified radiologists who

had undergone standardized training and had less than five years' experience in reading US images of renal diseases. The senior professional physician group consists of three professional radiologists with more than 10 years' experience in reading US images of renal diseases. The professionals read the US images in the test set and independently judged them as normal or diseased. Using the simple majority method, the diagnosis results of the two groups of professional doctors were determined for each image.

2.5 | Statistical analysis

Statistical analyses were performed using SPSS (version 25.0; IBM Corp., Armonk, NY, USA). The accuracy, sensitivity, specificity, and positive and negative predictive values of each group were calculated and compared. In order to compare the diagnostic efficacy of the screening model and the two groups of US physicians in diagnosing CKD, receiver operating characteristic (ROC) curves were used, and the area under the ROC curve (AUC) was determined. The AUCs were compared between the groups using the Mann–Whitney U test. All statistical analyses were two-sided with statistical significance set at $p < .05$.

3 | RESULTS

3.1 | Baseline characteristics

In total, 1049 patients were retrospectively enrolled, and a total of 4365 renal US images were collected from these patients. 756 US images in the non-CKD group were obtained from 179 patients, 597 US images in the G1-CKD group were obtained from 145 patients, 696 US images in the G2-CKD group were obtained from 169 patients, 1217 US images in the G3-CKD group were obtained from 289 patients, 536 US images in the G4-CKD group

were obtained from 129 patients, and 563 US images in the G5-CKD group were obtained from 138 patients. There were no significant differences in the average ages and sex ratios of patients in the training set, testing set, and verification set among each group (Table 1).

3.2 | Comparison of non-CKD diagnostic efficacy

In this study, the AUC, accuracy, sensitivity, and specificity of the junior radiologists group, senior radiologists group, and our model for the diagnosis of patients with non-CKD based on the characteristics of renal US images were 0.860, 70.9%, 72.9%, and 70.1% (junior group), 0.869, 72.3%, 74.7%, and 71.3% (senior group), and 0.918, 94.1%, 81.0%, and 96.4% (our model), respectively, as shown in Tables 2–4. In comparing the efficacy of the three groups in diagnosing non-CKD by AUC, there was no statistically significant difference in the diagnostic ability of senior and junior radiologists, $p > .05$ (AUC:0.860 vs. 0.869，$Z = 0.57$), as shown in Table 3. The diagnostic efficacy of our model was better than that of the senior radiologists group, $p < .001$ (AUC: 0.918 vs. 0.869，$Z = 7.16$), as shown in Table 4. The ROC curves are shown in Figure 3.

3.3 | Comparison of CKD stage diagnostic efficacy

For the diagnosis of CKD in the G1-G2 stage, although the sensitivity of the senior radiologists group was slightly better than that of the junior radiologists group, the diagnostic sensitivity of the senior radiologists group and junior radiologists group was less than 0.5. The diagnostic efficiency of the senior radiologists group and junior radiologists group was poor (G1-AUC: 0.506 vs. 0.483, G2-AUC: 0.586 vs. 0.533), with no statistically significant difference (G1: $Z = 0.74$, $p = .46$; G2: $Z = 1.76$, $p = .08$), as shown in Tables 2 and 3. The diagnostic sensitivity of our model exceeded 0.5 (G1: 60%, G2: 79.2%), and its diagnostic efficacy was significantly stronger than that

TABLE 1 Baseline characteristics of the participants in the different data sets.

		Training set	Test set	Validation set	Statistics	p value
Non-CKD	Age (year)	50.8 ± 19.3	49.3 ± 20.7	49.6 ± 19.9	F = .027	.973
	Gender (male/female)	77/66	10/8	11/7	$\chi^2 = .203$	.904
G1	Age (year)	52.5 ± 18.4	53.3 ± 19.7	52.6 ± 18.8	F = .02	.98
	Gender (male/female)	65/50	8/7	7/8	$\chi^2 = .204$	.903
G2	Age (year)	55.7 ± 17.6	56.3 ± 16.4	55.6 ± 18.1	F = .031	.91
	Gender (male/female)	70/65	10/7	9/8	$\chi^2 = .245$	.78
G3	Age (year)	58.1 ± 18.7	57.8 ± 19.1	58.3 ± 18.4	F = .035	.83
	Gender (male/female)	123/108	16/13	15/14	$\chi^2 = .214$	.87
G4	Age (year)	61.3 ± 19.5	61.9 ± 18.6	61.1 ± 19.7	F = .025	.977
	Gender (male/female)	57/46	8/5	7/6	$\chi^2 = .231$	.824
G5	Age (year)	65.3 ± 16.1	64.8 ± 17.7	65.4 ± 17.9	F = .033	.85
	Gender (male/female)	59/51	7/7	7/7	$\chi^2 = .313$	.634

TABLE 2 Comparison of the CKD diagnostic value of junior professionals, senior professionals, and the model.

Stage	Method	AUC	ACC(%)	SEN(%)	SPN(%)	PPV(%)	NPV(%)
Non-CKD	Junior	0.860	70.9	72.9	70.1	48.5	86.9
	Senior	0.869	72.3	74.7	71.3	49.3	88.3
	Our Model	0.918	94.1	81.0	96.4	80.3	96.5
G1	Junior	0.483	80.0	5.1	91.4	8.47	86.3
	Senior	0.506	81.7	8.3	92.9	15.1	86.9
	Our model	0.781	91.3	60.0	96.1	70.4	94.0
G2	Junior	0.533	79.0	15.5	91.0	24.7	85.0
	Senior	0.586	83.2	22.4	94.7	44.4	86.6
	Our model	0.880	94.0	79.2	96.8	82.7	96.0
G3	Junior	0.738	83.1	51.0	96.4	85.9	82.5
	Senior	0.796	87.0	61.8	97.5	91.1	85.9
	Our model	0.905	93.3	83.7	97.2	92.4	93.7
G4	Junior	0.857	93.4	75.8	95.7	69.1	96.9
	Senior	0.838	92.4	72.7	95.0	64.9	96.4
	Our model	0.867	94.6	76.5	97.0	76.5	97.0
G5	Junior	0.943	98.2	89.4	99.2	92.6	98.9
	Senior	0.963	97.8	94.3	98.2	84.7	99.3
	Our model	0.931	98.0	86.7	99.3	80.3	96.6

TABLE 3 Comparison of the CKD diagnostic ability of junior and senior professionals.

					95% confidence interval		U test	
	Test variables	AUC	Standard error	Significance	Lower bound	Upper bound	Z value	p value
Non-CKD	Junior	0.860	0.009	<.001	0.842	0.878	.57	.57
	Senior	0.869	0.009	<.001	0.851	0.886		
G1	Junior	0.483	0.022	.450	0.440	0.526	.74	.46
	Senior	0.506	0.022	.794	0.462	0.550		
G2	Junior	0.533	0.021	.114	0.491	0.574	1.76	.08
	Senior	0.586	0.022	<.001	0.542	0.629		
G3	Junior	0.738	0.016	<.001	0.706	0.769	2.65	<.01
	Senior	0.796	0.015	<.001	0.767	0.826		
G4	Junior	0.857	0.020	<.001	0.818	0.897	.66	.51
	Senior	0.838	0.021	<.001	0.797	0.880		
G5	Junior	0.943	0.015	<.001	0.913	0.973	1.08	.28
	Senior	0.963	0.011	<.001	0.940	0.985		

of the senior radiologists group (G1 AUC: 0.781 vs. 0.506, G2 AUC: 0.880 vs. 0.586), with statistically significant differences (G1: $Z = 8.84$, $p < .001$; G2: $Z = 10.81$, $p < .001$), as shown in Tables 2 and 4. The ROC curves are shown in Figure 3.

For the diagnosis of CKD in the G3 stage, the sensitivity of the junior radiologists group, senior radiologists group, and our model were 51.0%, 61.8%, and 83.7%, respectively. The detailed results are shown in Table 2. The diagnostic efficacy of the senior radiologists group was better than that of the junior radiologists group (AUC: 0.796 vs. 0.738), with a statistically significant difference ($Z = 2.65$, $p < .01$). Moreover, the diagnostic efficacy of our model was better

than that of the senior radiologists (AUC: 0.905 vs. 0.796), with a statistically significant difference ($Z = 5.86$, $p < .001$), as shown in Tables 3 and 4. The ROC curves are shown in Figure 3.

For the diagnosis of CKD in the G4 stage, the sensitivity of the junior radiologists group, senior radiologists group, and our model were 75.8%, 72.7%, and 76.5%, respectively. For the diagnosis of CKD in the G5 stage, the sensitivity of the junior radiologists group, senior radiologists group, and our model were 89.4%, 94.3%, and 86.7%, respectively. The detailed results are shown in Table 2. For the diagnostic efficacy of CKD in the G4-G5 stage, the diagnostic efficacy of the junior radiologists group, senior radiologists group,

TABLE 4 Comparison of the CKD diagnostic ability of senior professionals and our model.

	Test variables	AUC	Standard error	Significance	95% Confidence interval		U test	
					Lower bound	Upper bound	Z value	p value
Non-CKD	Senior	0.869	0.009	<.001	0.851	0.886	7.16	<.001
	Our model	0.918	0.016	<.001	0.896	0.939		
G1	Senior	0.506	0.022	.794	0.462	0.550	8.84	<.001
	Our model	0.781	0.022	<.001	0.737	0.824		
G2	Senior	0.586	0.022	<.001	0.542	0.629	10.81	<.001
	Our model	0.880	0.016	<.001	0.848	0.911		
G3	Senior	0.796	0.015	<.001	0.767	0.826	5.86	<.001
	Our model	0.905	0.011	<.001	0.883	0.927		
G4	Senior	0.838	0.021	<.001	0.797	0.880	1	.32
	Our model	0.867	0.020	<.001	0.829	0.906		
G5	Senior	0.963	0.011	<.001	0.940	0.985	1.65	.099
	Our model	0.931	0.016	<.001	0.899	0.963		
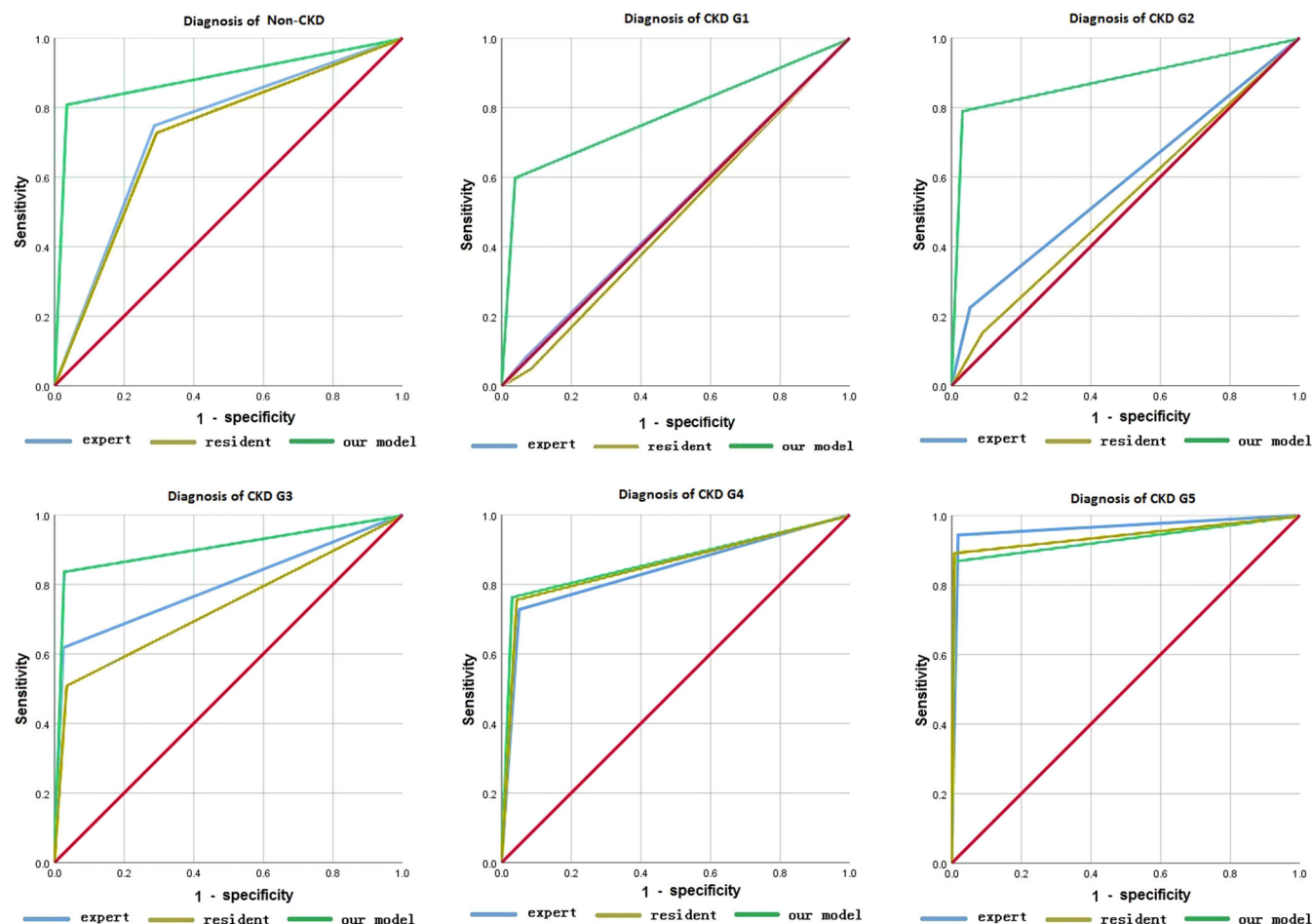
FIGURE 3 Receiver operating characteristic curves of the three groups for the diagnosis of CKD.

and our model were strong (G4-AUC: 0.838 vs. 0.857 vs. 0.867, G5-AUC: 0.963 vs. 0.943 vs. 0.931), and there was no significant difference between the junior radiologists group and senior radiologists group (G4: $Z = 0.66$, $p = .51$; G5: $Z = 1.08$, $p = .28$). There was also no statistically significant difference between our model and the senior radiologists group (G4: $Z = 1$, $p = .32$; G5:

$Z = 1.65$, $p = .099$), as shown in Tables 3 and 4. The ROC curves are shown in Figure 3.

4 | DISCUSSION

The commonly used imaging techniques for renal diseases mainly include x-ray, CT, MRI, and US. X-ray imaging and CT show the density differences between tissues and organs and between normal tissues and pathological tissues. Because of the low-density resolution of x-ray imaging, it can only show the outline of the kidney or stones in the kidney and is generally not used for the diagnosis and screening of CKD. CT imaging technology has a high-density resolution and can directly display the anatomical structures and pathological shapes that x-ray imaging cannot display. However, when diagnosing diseases, it is often necessary to obtain more blood flow and tissue information by injecting CT contrast enhancer. This approach has a good diagnostic effect for trauma, tumours, inflammation, congenital malformations, and other diseases. However, for patients with CKD, injecting CT contrast enhancer will increase the burden on renal function and even aggravate the disease. Considering radiation damage and renal function damage, x-ray and CT are not the first choice for the diagnosis and screening of CKD. MRI imaging can obtain rich renal image signal information by adjusting the parameters of the scanning sequence without injecting contrast enhancer. However, MRI examinations take a long time and are expensive. In addition, they cannot be used for some patients with metal implanted in the body. As such, MRI is not suitable as the first choice for the diagnosis and screening of CKD. US images are formed by using different echoes generated by attenuation differences in acoustic impedance between different tissues of the human body. US imaging has a high resolution and can display anatomical structures and pathological changes in the tissues. Compared with x-ray, CT, and MRI, US imaging has the advantages of real-time convenience, no radioactivity, no nephrotoxicity, and low cost. In addition, for the diagnosis of CKD, different from renal tumours, trauma, or inflammation, doctors pay more attention to the size of the kidneys, the blood flow, parenchyma thickness, and whether the boundary between the cortex and medulla is clear, all of which reflect the functional state of the kidneys. The ability to examine these specific markers makes US advantageous for CKD diagnosis. As such, US imaging is widely used for the prediction, diagnosis, and evaluation of CKD.[9,21] However, these applications are mostly limited to manual analysis using US image features and quantitative parameters, which rely on the personal experience of professional physicians and cannot be standardized. With the progression of CKD, US images will show an increasingly irregular kidney shape with a decreasing diameter, thinner renal cortex, unclear cortex-medulla boundary, and stronger echo.[22,23] In this study, the sensitivity of the senior radiologists group in the diagnosis of CKD according to renal US images was 74.7%, while that of the junior radiologists group was 72.9%. In the staging diagnosis of CKD, although the diagnostic sensitivity of the senior radiologists group and junior radiologists group gradually increased with increasing

staging degree of CKD, the diagnostic sensitivity of the senior radiologists group for CKD G1–G4 was only 8.3%, 22.4%, 61.8%, and 72.7% respectively. Therefore, it can be concluded that it is very difficult for professional radiologists to diagnose all except the most severe stages of CKD by reading US images.

There has been rapid development of AI and imageomics technology (IT) in recent years. To improve radiologists' judgement of the severity of CKD, many researchers have proposed various types of ancillary diagnostic methods to predict CKD staging[24] with the use of machine learning methods based on clinical data such as age, weight, blood pressure, and blood glucose, and the use of classification algorithms involving artificial neural networks (ANN),[25] Naive Bayes (NB),[26] the K-nearest neighbour algorithm (KNN),[27] and support vector machine (SVM).[28] Although these methods have achieved high classification accuracy, they require the collection of a wide variety of clinical data, which is not always available. Therefore, many researchers have applied deep learning techniques based on IT to diagnose various kidney diseases, including renal cysts, renal stones, renal tumours, etc.[29,30] However, research inadequacy and a gap in CKD-related radiomics studies still remain. Li Guanghan and colleagues applied a deep learning model based on the convolutional neural network structure DenseNet121 to the US diagnosis of end-stage renal disease and found that compared with manual reading by professional doctors, the deep learning model improved the detection rate of end-stage renal disease.[30]

There exist similarities between normal kidney US images and CKD US images, and there are also differences among samples in each stage of CKD. Therefore, compared to the classification of natural images, the diagnosis of CKD is much more difficult. This further increases the difficulty of using deep learning methods to achieve high accuracy in this task. The base network, as the backbone of the model, should be chosen with great care. In spite of the fact that more complex networks have stronger feature expression abilities, the Residual Network (ResNet) with moderate complexity was selected in this study because the number of CKD samples was small, and a complex network structure is more likely to fall into overfitting at the training stage. After comparison, ResNet 34 was chosen as the base network for the current model. As an attempt to integrate deep learning into CKD evaluation, this study found that the model constructed by fusing ResNet34 and texture features had noticeable efficiency for the diagnosis of CKD at all stages of the disease. When diagnosing CKD or non-CKD based on renal US images, the results of this study showed that the accuracy, sensitivity, specificity, and AUC of our model were 21.8%, 6.3%, 25.1%, and 0.05 higher than those of the senior radiologists group. Moreover, the diagnostic efficiency was higher than that of the senior radiologists group, and the difference in the AUC between our model and the senior radiologists group was statistically significant. For the diagnosis of CKD stage, the diagnostic sensitivity of our model for CKD G1-G3 was significantly higher than that of the senior radiologists group, with increases of 51.7%, 56.8%, and 21.9% respectively, and the diagnostic efficiency was also significantly higher than that of the senior radiologists group. The difference in the AUCs was statistically significant. The diagnostic efficiency of

our model for CKD G4 and G5 reached the level of the senior radiologists group, with no statistically significant difference in the AUC between our model and the senior radiologists group.

This study offers valid evidence that our model outperforms manual image reading, with higher accuracy and sensitivity. The model can effectively improve the detection rate of CKD in patients who have not undergone biochemical tests and can, therefore, guide further clinical examinations and treatment. However, there are still several limitations of this study. First, the measured values of the renal US images of people in different countries and regions are different.[31] Further, due to differences in hardware and software parameters, the image information extracted from renal US images collected by ultrasound machines from different manufacturers will differ. However, this study was only a single-center, small-sample retrospective study, and the collected US image information was relatively simple. Further, the number of US images for the different stages of CKD was not balanced, and thus, the established model will be difficult to generalize. In the future, we plan to use a variety of different ultrasonic equipment to collect images from CKD patients in multiple hospitals in multiple regions and conduct a large-sample multicenter study to address this problem. Secondly, a comprehensive and credible prediction model needs not only internal verification but also external verification. External verification not only provides an evaluation of the performance of the model but also helps to prevent over-fitting, optimize the model parameters, select the best model, and enhance the generalization ability of the model. Because this research was only preliminary, we stipulated and required the operators' experience and ultrasonic equipment model when establishing the model and did not examine an external verification set. In the future, we will perform external verification while conducting large-sample multi-center research to expand the audience and further determine the evaluation effect of the prediction model. Finally, although our model performed well in this study, only the ultrasonic image information of the renal contour and renal cortex was extracted; information on the renal medulla, renal pelvis, and CKD-related blood laboratory indices was ignored. Thus, follow-up research is needed to further explore more comprehensive information from US images and combine this with an examination of other clinical factors closely related to CKD to build a more robust and comprehensive model, so as to further improve the diagnostic performance of the model.

Based on our preliminary research results, we believe that we can establish a CKD diagnosis model with good generalization ability and stronger robustness after scientifically solving the limitations of the current research. With a more powerful model, we plan to further build an interactive CKD screening system based on the Python and Django frameworks, which can be easily accessed through an internet browser. We will also integrate this into the PACS system of the imaging department so that when the imaging doctor writes the renal US image report, the patient can be diagnosed and graded through our screening model. With such a CKD screening system, patients can be immediately warned whether they have CKD and its degree through ultrasonic images. In addition, this model can be used as a tool to monitor the progress of patients with CKD. This will reduce the number of times the patient needs to have repeated checks of renal function, electrolytes, and other items, which will save medical resources. In addition, as a reliable new technology for diagnosing CKD, this model can help enhance the competitiveness of medical institutions and attract more patients, thus offering economic benefits to medical institutions.

In conclusion, features based on grey-scale US, as a quantitative analysis method, can provide richer information for the diagnosis of CKD. The model constructed here based on these features had good performance for CKD screening and stage diagnosis. It is anticipated that the model can assist radiologists in diagnosing CKD while reducing the number of missed diagnoses and avoiding invasive tests. The findings of this study should be verified in future studies with larger sample sizes.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ORCID

Shuyuan Tian https://orcid.org/0000-0001-5148-3733

REFERENCES

1. Alnazer I, Bourdon P, Urruty T, et al. Recent advances in medical image processing for the evaluation of chronic kidney disease. *Med Image Anal*. 2021;69:101960.
2. Zhang L, Zhao MH, Zuo L, et al. China kidney disease network (CK-NET) 2016 annual data report. *Kidney Int Suppl*. 2011;10(2):e97-e185.
3. Bikbov B, Purcell CA, Levey AS, et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet*. 2020; 395(10225):709-733.
4. Legrand K, Speyer E, Stengel B, et al. Perceived health and Qualityof life in patients with CKD, including those with kidney failure: findings from National Surveys in France. *Am J Kidney Dis*. 2020;75(6): 868-878.
5. Benito S, Sanchez-Ortega A, Unceta N, et al. LC-QQQ-MS routine analysis method for new biomarker quantification in plasma aimed at early chronic kidney disease diagnosis. *J Pharm Biomed Anal*. 2019; 169(1):82-89.
6. Ikizler TA, Burrowes JD, Byham-Gray LD, et al. KDOQI clinical practice guideline for nutrition in CKD: 2020 update. *Am J Kidney Dis*. 2020;76(3 Suppl 1):S1-S107.
7. Fan LL, Liu L, Luo FM, et al. A novel heterozygous variant of theCOL4A4gene in a Chinese family with hematuria and proteinuria leads to

focal segmental glomerulosclerosis and chronic kidney disease. *Mol Genet Genomic Med*. 2020;8(12):e1545.

8. Halimi JM, Gatault P, Longuet H, et al. Major bleeding and risk of death after percutaneous native kidney biopsies: a French Nationwide cohort study. *Clin J Am Soc Nephrol*. 2020;15(11):1587-1594.

9. Petrucci I, Clementi A, Sessa C, Torrisi I, Meola M. Ultrasound and color Doppler applications in chronic kidney disease. *J Nephrol*. 2018; 31(6):863-879.

10. Kuo CC, Chang CM, Liu KT, et al. Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. *NPJ Digit Med*. 2019;2:29.

11. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging "how-to" guide and critical reflection. *Insights Imaging*. 2020;11(1):91.

12. Gillies RJ, Schabath MB. Radiomics improves cancer screening and early detection. *Cancer Epidemiol Biomarkers Prev*. 2020;29(12):2556-2567.

13. Gu J, Tong T, He C, et al. Deep learning radiomics of ultrasonography can predict response to neoadjuvant chemotherapy in breast cancer at an early stage of treatment: a prospective study. *Eur Radiol*. 2021; 32(3):2099-2109.

14. Yu B, Li Y, Yu X, et al. Differentiate thyroid follicular adenoma from carcinoma with combined ultrasound Radiomics features and clinical ultrasound features. *J Digit Imaging*. 2022;35(5):1362-1372.

15. Xiong L, Chen H, Tang X, et al. Ultrasound-based Radiomics analysis for predicting disease-free survival of invasive breast cancer. *Front Oncol*. 2021;11:621993.

16. Wu JP, Ding WZ, Wang YL, et al. Radiomics analysis of ultrasound to predict recurrence of hepatocellular carcinoma after microwave ablation. *Int J Hyperth*. 2022;39(1):595-604.

17. Roberts MA. Commentary on the KDIGO clinical practice guideline for the management of blood pressure in chronic kidney disease. *Nephrol Ther*. 2014;19(1):53-55.

18. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *Systems Man & Cybernetics IEEE Transactions*. 1973; SMC-3(6):610-621.

19. Liu C, Hu Y, Chen Y, et al. Improvements to a GLCM-based machine-learning approach for quantifying posterior capsule opacification. *J Appl Clin Med Phys*. 2024;25:e14268. doi:10.1002/acm2.14268

20. Dalal N, Triggs B. Histograms of oriented gradients for human detection. *Conference on Computer Vision & Pattern Recognition*. IEEE; 2005.

21. Xie TS, Zhou ZR. The progress and application of artificial intelligence and radiomics on abdominal tumors. *Chin J Radiol*. 2020;54(4):376-379.

22. Guo HY. The research progress on ultrasound evaluation of chronic kidney disease. *Henan Med Res*. 2013;04:638-640.

23. Liu BY, Zhao LX, Zheng SG, et al. Ultrasound multi-parameter scoring in diagnosis of chronic kidney disease. *Chin J Med Imag Technol*. 2021; 02:273-277.

24. Batra A, Batra U, Singh V. A review to predictive methodology to diagnose chronic kidney disease. International Conference on Computing for Sustainable Global Development. 2760–2763 2016.

25. Jeewantha RA, Halgamuge MN, Mohammad A, et al. Classification performance analysis in medical science: using kidney disease data. International Conference on Big Data Research. 2017.

26. Kunwar V, Chandel K, Sabitha AS, et al. Chronic kidney disease analysis using data mining classification techniques. International Conference on Cloud System and Big Data Engineering, 300–305. 2016.

27. Charleonnan A, Fufaung T, Niyomwong T, et al. Predictive analytics for chronic kidney disease using machine learning techniques. Management and Innovation Technology International Conference. 2016.

28. Ahmad M, Tundjungsari V, Widianti D, et al. Diagnostic Decision Support System of Chronic Kidney Disease Using Support Vector Machine. International Conference on Informatics and Computing. 2017.

29. Wu Y, Yi Z. Automated detection of kidney abnormalities using multi-feature fusion convolutional neural networks. *Knowl-Based Syst*. 2020;200(2):105873.

30. Li GH, Liu J, Ma LY, et al. Deep learning-based model for ultrasound diagnosis of end-stage chronic kidney disease. *Chin J Med Ultrasound (Electronic Edition)*. 2021;18(6):611-615.

31. Ahmad RU. *Evaluation of Normal Renal Values by Ultrasound in Chinese and Pakistani Population and their Comparison*. Shandong University; 2013.