Artificial intelligence for predicting interstitial fibrosis and tubular atrophy using diagnostic ultrasound imaging and biomarkers

Ting-Wei Chang ,¹ Chang-Yu Tsai,² Zhen-Yi Tang,² Cai-Mei Zheng,^{3,4,5} Chia-Te Liao,^{3,4,5} Chung-Yi Cheng,^{3,4,5} Mai-Szu Wu,^{3,4,5} Che-Chou Shen,² Yen-Chung Lin

ABSTRACT

Background Chronic kidney disease (CKD) is a global health concern characterised by irreversible renal damage that is often assessed using invasive renal biopsy. Accurate evaluation of interstitial fibrosis and tubular atrophy (IFTA) is crucial for CKD management. This study aimed to leverage machine learning (ML) models to predict IFTA using a combination of ultrasonography (US) images and patient biomarkers.

Methods We retrospectively collected US images and biomarkers from 632 patients with CKD across three hospitals. The data were subjected to pre-processing, exclusion of sub-optimal images, and feature extraction using a dual-path convolutional neural network. Various ML models, including XGBoost, random forest and logistic regression, were trained and validated using fivefold cross-validation.

Results The dataset was divided into training and test datasets. For image-level IFTA classification, the best performance was achieved by combining US image features and patient biomarkers, with logistic regression yielding an area under the receiver operating characteristic curve (AUROC) of 99%. At the patient level, logistic regression combining US image features and biomarkers provided an AUROC of 96%. Models trained solely on US image features or biomarkers also exhibited high performance, with AUROC exceeding 80%.

Conclusion Our artificial intelligence-based approach to IFTA classification demonstrated high accuracy and AUROC across various ML models. By leveraging patient biomarkers alone, this method offers a non-invasive and robust tool for early CKD assessment, demonstrating that biomarkers alone may suffice for accurate predictions without the added complexity of image-derived features.

INTRODUCTION

Chronic kidney disease (CKD) causes significant morbidity and mortality worldwide, with global prevalence rates of 9.1% and 697.5 million cases.¹ It is characterised by irreversible damage to the renal tissue, which can ultimately lead to end-stage kidney disease, resulting in a substantial economic burden.²

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Chronic kidney disease (CKD) is a significant global health issue, with accurate evaluation of interstitial fibrosis and tubular atrophy (IFTA) being essential for its management, typically requiring invasive renal biopsy.

WHAT THIS STUDY ADDS

⇒ This study demonstrates that combining ultrasonography (US) images and patient biomarkers using machine learning (ML) models can accurately predict IFTA non-invasively, achieving high area under the receiver operating characteristic curve values with logistic regression models.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The findings suggest that an ML-based approach integrating US images and biomarkers can serve as a non-invasive, reliable tool for early CKD assessment, potentially enhancing clinical decision-making and patient outcomes while reducing the need for invasive procedures.

A previous study showed that accurate assessment of renal interstitial fibrosis and tubular atrophy (IFTA) is crucial for diagnosing and managing CKD.³ IFTA severity is conventionally assessed through renal biopsy, which remains the gold standard for obtaining detailed histopathological information. This procedure provides direct visualisation and quantification of IFTA but is time-consuming and subject to inter-observer variability. Additionally, renal biopsy is an invasive procedure, making it unsuitable for all patients.⁴

Medical ultrasonography (US) is a crucial diagnostic tool for kidney or ureteral structural diseases, which measures imaging parameters that provide vital information regarding renal function. Previous studies have explored different parameters such as

To cite: Chang T-W, Tsai C-Y, Tang Z-Y, *et al.* Artificial intelligence for predicting interstitial fibrosis and tubular atrophy using diagnostic ultrasound imaging and biomarkers. *BMJ Health Care Inform* 2025;**32**:e101192. doi:10.1136/bmjhci-2024-101192

Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/ bmjhci-2024-101192).

Received 08 July 2024 Accepted 02 March 2025

Check for updates

© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to Dr Yen-Chung Lin; yclin0229@tmu.edu.tw

5 1

kidney size, cortical thickness and cortical echogenicity to estimate changes in estimated glomerular filtration rate (eGFR).^{5–12} A previous study revealed a significant positive correlation between eGFR and the mean renal length (r=0.66) and mean cortical thickness (r=0.85).⁵ Additionally, changes in the texture of the renal tissue on US images can also suggest changes in renal function. Nevertheless, interpreting US images of the kidney requires extensive training for clinicians, and the results may lack objectivity due to subjectivity.

Recently, artificial intelligence (AI) has emerged as a promising tool for predicting pathological results by leveraging data acquired through non-invasive methods.¹³¹⁴ Deep learning using convolutional neural networks (CNNs) has also demonstrated good performance in the analysis of medical US images. In a previous study, CNNs were used to grade the severity of inflammation in the long head of the bicep tendon.¹⁵ Moreover, one study relied on CNNs and consecutive comprehensive non-stress echocardiography to predict cardiac function in patients to better understand ageing and prevent cardiovascular diseases.¹⁶ Similarly, a deep-learning algorithm using kidney US images of a single centre accurately quantified IFTA with 90% accuracy, indicating its potential as a non-invasive first-line investigation for kidney-disease assessment.¹⁷ However, there is a lack of studies conducted across multiple medical centres in leveraging AI for predicting pathological results through non-invasive data acquisition methods.

Our research aimed to leverage AI to predict the stage of IFTA using a combination of demographic data, laboratory results and renal US images across diverse medical centres. Specifically, we focused on five key clinical biomarkers: age, sex, eGFR, serum albumin and kidney size. Renal US provides a non-invasive and readily available imaging modality that can capture detailed structural information regarding the kidneys. We compared three approaches: using only biomarkers, relying solely on US images and integrating both modalities for a comprehensive analysis.

METHODS

Datasets

The dataset used in this study was retrospectively obtained from Taipei Medical University Hospital, Taipei Municipal Wanfang Hospital and Taipei Medical University Shuang Ho Hospital. Patients who underwent pre-biopsy US examination in the past 10 years were included in the study. Renal ultrasound images obtained within 3 months before or after the renal biopsy, and the renal pathology reports closest to the biopsy date, were collected for each patient. The Joint Institutional Review Board Committee of Taipei Medical University (TMU-JIRB) approved the study (no. N202008034). Patient consent was waived by the permit of TMU-JIRB due to practical problems that most of the patients were lost to follow-up.

Patient selection

The data on key patient biomarkers including creatinine levels, age, sex and urine protein levels were extracted from our database. Additionally, the kidney size was obtained from US reports. We excluded images of patients with diabetic nephropathy to prevent inaccurate identification of US image features due to enlarged kidneys.^{18 19} The IFTA can be stratified into four individual stages.²⁰ A previous study showed that an IFTA of >25% is an indicator of progression to endstage renal disease.²¹ Hence, we used the combined IFTA stages to express the degree of renal damage, which ranged from 0 to 3. We divided the data into two groups: scores 0 and 1 were classified as the mild group and scores 2 and 3 were classified as the severe group, making the dataset a binary classification. A total of 632 patients with 6029 images were included.

Development of IFTA-classification pipeline

The development of the AI model involved four independent steps: (1) US image pre-processing, (2) sub-optimal image exclusion, (3) US image feature extraction and (4) ML-based IFTA classification with internal and independent validation. The detailed pipeline is shown in figure 1.

US image pre-processing

We used a Mask R-CNN to extract the kidney region to eliminate background interference from the muscles, adipose tissue, liver, spleen and bowel during IFTA prediction.²² We manually marked the kidney region in 2000 images (1800 for training and 200 for testing) for Mask R-CNN training.

Sub-optimal image or incomplete data exclusion

Based on previous studies,¹⁸ ¹⁹ we excluded all US images from patients diagnosed with diabetic nephropathy because this condition can cause the kidneys to appear unusually large and functionally enhanced, potentially misleading our image feature extractor. Additionally, we removed images with significant artefacts such as prominent acoustic shadows or incorrect segmentations, as well as those in which the kidneys were not completely captured. Patients with incomplete IFTA data were excluded. After these exclusions, the remaining dataset consisted of 1895 images from 266 patients.

US image features extractor

This training was conducted at the image level after data augmentation. The augmentation methods included rotation between -25° and 25°, horizontal flip, vertical shift, gamma correction, Gaussian blur, brightness adjustment and contrast-limited adaptive histogram equalisation,²³ expanding the image dataset to 36 339 images. We trained a classification model using a dual-path convolutional neural network (DPCNN) on US images, which was well suited for US image classification tasks in a previous study.²⁴ The



Figure 1 Overall classification pipeline F2048 represents a 2048-dimension feature vector from the feature extractor and BM5 represents five key biomarkers. DPCNN, dual-path convolutional neural network; IFTA, interstitial fibrosis and tubular atrophy; ML, machine learning; ROI, region of interest.

trained DPCNN model was then used as a feature extractor to generate a 2048-dimensional vector for each US image.

IFTA severity classification

ML-based IFTA classification

To classify the severity of the IFTA, we employed advanced ML models such as extreme gradient boosting (XGBoost),²⁵ random forest²⁶ and light gradient boosting machine (LightGBM).²⁷ Ensemble methods were also implemented. We used a greedy search to find the best ML model and deployed fivefold cross-validation and averaged the evaluation results on the test set.

These models were trained on three different types of input:

- 1. Patient biomarkers: the data on five key patient biomarkers, including age, sex, eGFR, serum albumin and kidney size from US reports.
- 2. US image features: the 2048-dimensional vectors were extracted from the US images using the trained DPCNN model.

3. Combined inputs: a combination of five patient biomarkers and US image features were employed.

By leveraging these diverse inputs, we aimed to enhance the performance and robustness of the IFTA-severity classification. The scripts were implemented using Python 3.6 (Python Software Foundation) and executed on a system equipped with an RTX 2080 Ti Graphics Processing Unit.

Statistical analysis

To assess the differences in biomarkers among groups with different IFTA severity levels, we used independent t-test for continuous variables and two-sided Pearson χ^2 test for categorical variables. The model performance was evaluated using several metrics: accuracy, precision, recall (sensitivity), F1 score (calculated as the harmonic mean of precision and recall) and area under the receiver operating characteristic curve (AUROC), which measures the ability of the model to discriminate between classes across all threshold values. To compare the AUROC between different models or groups, we applied the DeLong test,²⁸ a non-parametric method, to assess the

	Participant, no. (%)	Participant, no. (%)		
	IFTA 0%–25%	IFTA >25%		
haracteristic	(n=129)	(n=94)	P value	
ge, mean (SD), years	48.2 (16.1)	53.1 (16.2)	0.025	
ex				
Female	59 (45.7)	43 (45.7)	1	
Male	70 (54.3)	51 (54.3)		
GFR (mL/min/1.73 m ²)	74.6 (39.7)	30.4 (25.2)	<0.001	
erum albumin (g/dL)	3.3 (1.1)	3.6 (0.7)	0.011	
dney size (cm)	10.5 (1.1)	9.9 (1.3)	< 0.001	

eGFR, estimated glomerular filtration rate; IFTA, interstitial fibrosis and tubular atrophy.

statistical significance of the difference between two AUROC values.

RESULTS

Study participants

A total of 266 patients were included in the study after applying the aforementioned exclusion criteria. We initially partitioned 60 patients with complete US images to serve as an independent test set to validate subsequent models. The data of the remaining 206 patients were used to train the image feature extractor. Patients with incomplete biomarkers were excluded from both the training and test sets. After the exclusion, 171 and 52 patients with complete US images and biomarker data were respectively used to train and to test the ML models. Detailed statistical information of the dataset for ML models is provided in table 1.

Classification using eGFR

We used eGFR alone as the baseline classification. The eGFR values were normalised and used for filtering. The optimal threshold for eGFR was determined by identifying the value that maximised accuracy. The threshold of $40.7 \,\mathrm{mL/min}/1.73 \,\mathrm{m^2}$ yielded the highest accuracy at 0.865. The confusion matrix and AUROC for using eGFR as the threshold are illustrated in figure 2. These results served as a baseline for further analyses.

Kidney Segmentation

Mask R-CNN model was used to perform kidney segmentation using a dataset of 200 images. For images containing intact kidneys, the model demonstrated performance comparable to that of manual annotation. Specifically,



Figure 2 (A) The confusion matrix of using eGFR to classify IFTA. **B**. The AUROC plot of using eGFR to classify IFTA. AUROC, area under the receiver operating characteristic curve; eGFR, estimated glomerular filtration rate; IFTA, interstitial fibrosis and tubular atrophy.

Table 2	Image-level	and	patient-level	evaluation	metrics of	of XGBoost
---------	-------------	-----	---------------	------------	------------	------------

•	•					
	Image level		Patient level			
	F2048+BM5	F2048	F2048+BM5	F2048	BM5	
Accuracy	0.89	0.76	0.90	0.79	0.83	
Precision	0.94	0.72	0.96	0.80	0.90	
Recall	0.76	0.63	0.85	0.77	0.73	
F1 score	0.84	0.67	0.90	0.78	0.81	
AUROC	0.95	0.91	0.93	0.84	0.93	
P value	Ref	<0.01	Ref	0.07	0.92	

F2048 represents a 2048-dimension feature vector from the feature extractor and BM5 represents five key biomarkers. The p value denotes the result of the DeLong test, which compares the performance differences between different feature sets.

AUROC, area under the receiver operating characteristic curve.

the model achieved an Intersection over Union of 0.904 and a Dice coefficient of 0.949, indicating high accuracy and reliability in segmenting kidney structures.

Feature extraction

For feature extraction, the DPCNN was trained on a dataset of 1456 images over 100 epochs, and the model with the lowest validation loss was selected as the feature extractor. This approach ensured that the extracted features were robust and representative of the underlying data.

Image-level IFTA classification

When using only US image features, the XGBoost model achieved the best performance with an accuracy of 76%, a precision of 72%, a recall of 63%, an F1 score of 67% and an AUROC of 91%. Combining US image features with patient biomarkers, the logistic regression model performed best, yielding an accuracy of 93%, a precision of 98%, a recall of 85%, an F1 score of 91% and an AUROC of 99%. Detailed results are presented in tables 2 and 3 and online supplemental tables 1–9. Furthermore, based on the DeLong test, the models using the combined image and biomarker features showed statistically significant improvements over those using only image features.

Patient-level IFTA classification

For patient-level classification, the predicted classes of all images from the same patient were averaged with a threshold of 0.5 used to determine the final classification. Using US image features alone, the logistic regression model produced the best results, with an accuracy of 81%, a precision of 100%, a recall of 62%, an F1 score of 76% and an AUROC of 93%. Using patient biomarkers alone, the LightGBM model achieved the highest performance, with an accuracy of 90%, a precision of 100%, a recall of 81%, an F1 score of 89%, and an AUROC of 95% and AUROC of 94.97%. When combining both US image features and patient biomarkers, the logistic regression model provided the best results, with an accuracy of 88%, a precision of 92%, a recall of 85%, an F1 score of 88% and an AUROC of 96%. Based on the DeLong test, the models combining image features with biomarkers did not show statistically significant improvements over those using biomarkers alone.

DISCUSSION

The development and validation of our AI algorithm involved a multistep process to ensure robustness and accuracy for both image-level and patient-level classifications.

Table 3 Image-level and patient-level evaluation metrics of logistic regression.						
	Image level		Patient level			
	F2048+BM5	F2048	F2048+BM5	F2048	BM5	
Accuracy	0.93	0.78	0.88	0.81	0.87	
Precision	0.98	0.84	0.92	1	0.88	
Recall	0.85	0.54	0.85	0.62	0.85	
F1 score	0.91	0.66	0.88	0.76	0.86	
AUROC	0.99	0.86	0.96	0.93	0.92	
P value	Ref	<0.01	Ref	0.60	0.07	

F2048 represents a 2048-dimension feature vector from the feature extractor and BM5 represents five key biomarkers. The p value denotes the result of the DeLong test, which compares the performance differences between different feature sets. AUROC, area under the receiver operating characteristic curve.

Open access

The use of biomarkers alone yielded optimal results in our analyses, demonstrating that the structural insights from imaging did not significantly enhance predictive accuracy. The use of fivefold cross-validation helped minimise bias and optimise model performance, leading to high accuracy and AUROC scores across various classifiers.

Previous studies investigated the correlation between renal US images and IFTA scores.²⁹ These studies found that sonographic parameters such as kidney length, echogenicity and parenchymal thickness demonstrated only weak-to-moderate relationships with interstitial fibrosis or tubular atrophy, with the highest Spearman correlation coefficient reaching 0.35. In another study, researchers proposed an AI model to predict IFTA scores based on renal US images.¹⁷ The UNet architecture was employed for US image segmentation, and the VGG19 and XGBoost models were used for feature extraction and image classification, respectively. The prediction model achieved a performance ranging from 0.8037 to 0.8927 in terms of accuracy, precision, recall and F1 score. However, this study relied on retrospective data from a single centre.

One of the major strengths of our study is the comprehensive dataset, which included a large number of images from a diverse patient cohort across multiple hospitals, enhancing the generalisability of our model. Additionally, our approach combines the analysis of US images with patient biomarkers to provide a valuable layer of information that improves the predictive power of our models. This method demonstrated a robust performance across various ML techniques, consistently achieving an AUROC >90%. While it cannot entirely substitute human interpretation and clinical expertise, it can serve as a valuable tool for assisting less-experienced physicians in providing accurate interpretations. Future research should focus on investigating the long-term impact and cost-effectiveness of this technology, as well as exploring effective ways to integrate it into existing clinical practices.

Despite these promising results, our study has several limitations. First, the retrospective nature of the study may have introduced biases related to data collection and patient selection, which may have affected the generalisability of our findings. Second, the exclusion of patients with diabetic nephropathy, while necessary to avoid misleading the model, may limit the applicability of our results to a broader population, particularly those with comorbid conditions. Third, the manual review process for image quality is subject to human error and variability, which can affect the consistency of the data used for training and validation. Fourth, while our model demonstrated high performance, with AUROC >90% in the independent test set, it needs to be validated in external cohorts to confirm its robustness and generalisability beyond the studied population. Fifth, the lack of standardised pathology guidelines across hospitals in our country posed a significant limitation. As each pathology department follows different guidelines, detailed parameters such as IFTA foci density could not be consistently obtained or included in this study. Finally, follow-up

data were not available in our study to assess the model's capacity to predict long-term outcomes such as end-stage kidney disease (ESKD). Future studies should include follow-up information to evaluate the model's potential for predicting ESKD and its broader clinical utility.

CONCLUSIONS

Our study presents a robust AI-based approach for IFTA classification that relies on patient biomarkers alone, demonstrating that image features do not significantly enhance predictive performance. The model demonstrated high performance across various ML techniques, consistently achieving AUROC >90%. This method, based on biomarkers alone and developed using a comprehensive dataset from multiple hospitals, has the potential to enhance early determination of IFTA without biopsy, avoiding unnecessary complexity from image features. Future studies should focus on validating the performance of the model in external cohorts to ensure its generalisability and applicability to diverse clinical environments.

Author affiliations

¹Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

²Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

³Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

⁴TMU Research Center of Urology and Kidney (TMU-RCUK), Taipei Medical University, Taipei, Taiwan

⁵Division of Nephrology, Department of Internal Medicine, Shuang Ho Hospital, Taipei Medical University, Taipei, Taiwan

Acknowledgements The authors extend their appreciation to the National Science and Technology Council of Taiwan for funding and supporting this work through Grant No. 111-2221-E-011-020-MY2 and 113-2221-E-011-021-MY3, the National Taiwan University of Science and Technology, and Taipei Medical University through Grant No. TMU-NTUST-111-08.

Contributors YCL is responsible for the overall content (as guarantor). Conceptualisation: YCL and CCS. Data curation: CMZ, CYC, YCL and CCS. Methodology: TWC, CYT, YCL and CCS. Formal analysis: CYT, ZYT and CCS. Funding acquisition: YCL and CCS. Supervision: YCL, CTL, MSW and CCS. Writing—original draft: TWC. Writing—review and editing: TWC, CYT, YCL and CCS. No, I have not used AI.

Funding Funded by the National Science and Technology Council of Taiwan, supporting this work through Grant No. 111-2221-E-011-020-MY2 and 113-2221-E-011-021-MY3, the National Taiwan University of Science and Technology, and Taipei Medical University through Grant No. TMU-NTUST-111-08.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The Joint Institutional Review Board Committee of Taipei Medical University (TMU-JIRB) approved the study (no. N202008034).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines,

terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs

Ting-Wei Chang http://orcid.org/0009-0005-2243-1084 Yen-Chung Lin http://orcid.org/0000-0002-8981-8671

REFERENCES

- 1 Carney EF. The impact of chronic kidney disease on global health. *Nat Rev Nephrol* 2020;16:251.
- 2 Wang V, Vilme H, Maciejewski ML, *et al.* The Economic Burden of Chronic Kidney Disease and End-Stage Renal Disease. *Semin Nephrol* 2016;36:319–30.
- 3 Risdon RA, Sloper JC, De Wardener HE. Relationship between renal function and histological changes found in renal-biopsy specimens from patients with persistent glomerular nephritis. *Lancet* 1968;2:363–6.
- 4 Visconti L, Cernaro V, Ricciardi CA, et al. Renal biopsy: Still a landmark for the nephrologist. World J Nephrol 2016;5:321-7.
- 5 Lucisano G, Comi N, Pelagi E, et al. Can renal sonography be a reliable diagnostic tool in the assessment of chronic kidney disease? J Ultrasound Med 2015;34:299–306.
- 6 Korkmaz M, Aras B, Güneyli S, *et al.* Clinical significance of renal cortical thickness in patients with chronic kidney disease. *Ultrasonography* 2018;37:50–4.
- 7 Mustafiz M, Rahman MM, Islam MS, et al. Correlation of ultrasonographically determined renal cortical thickness and renal length with estimated glomerular filtration rate in chronic kidney disease patients. Bangladesh Med Res Counc Bull 2013;39:91–2.
- 8 Mansoor A, Ramzan A. Evaluation of the Best Gray-Scale Ultrasonography Parameter for Assessment of Renal Function in CKD Patients: A Meta-Analysis of Literature Published in the Last 10 years. *European Congress of Radiology-ECR* 2015.
- 9 Takata T, Koda M, Sugihara T, et al. Left Renal Cortical Thickness Measured by Ultrasound Can Predict Early Progression of Chronic Kidney Disease. Nephron 2016;132:25–32.
- 10 Beland MD, Walle NL, Machan JT, et al. Renal cortical thickness measured at ultrasound: is it better than renal length as an indicator of renal function in chronic kidney disease? AJR Am J Roentgenol 2010;195:W146–9.
- 11 Singh A, Gupta K, Chander R, *et al.* Sonographic grading of renal cortical echogenicity and raised serum creatinine in patients with chronic kidney disease. *Jemds* 2016;5:2279–86.

- 12 Yamashita SR, Atzingen AC von, lared W, et al. Value of renal cortical thickness as a predictor of renal function impairment in chronic renal disease patients. *Radiol Bras* 2015;48:12–6.
- 13 Liang L, Han X, Zhou N, et al. Ultrasound for Preoperatively Predicting Pathology Grade, Complete Cytoreduction Possibility, and Survival Outcomes of Pseudomyxoma Peritonei. Front Oncol 2021;11:690178.
- 14 Zhang Y, Qu H, Tian Y, *et al.* PB-LNet: a model for predicting pathological subtypes of pulmonary nodules on CT images. *BMC Cancer* 2023;23:936.
- 15 Lin B-S, Chen J-L, Tu Y-H, et al. Using Deep Learning in Ultrasound Imaging of Bicipital Peritendinous Effusion to Grade Inflammation Severity. IEEE J Biomed Health Inform 2020;24:1037–45.
- 16 Ghorbani A, Ouyang D, Abid A, et al. Deep learning interpretation of echocardiograms. NPJ Digit Med 2020;3:10.
- 17 Athavale AM, Hart PD, Itteera M, et al. Development and Validation of a Deep Learning Model to Quantify Interstitial Fibrosis and Tubular Atrophy From Kidney Ultrasonography Images. JAMA Netw Open 2021;4:e2111176.
- 18 Derchi LE, Martinoli C, Saffioti S, et al. Ultrasonographic imaging and Doppler analysis of renal changes in non-insulin-dependent diabetes mellitus. Acad Radiol 1994;1:100–5.
- 19 Buturović-Ponikvar J, Visnar-Perovic A. Ultrasonography in chronic renal failure. *Eur J Radiol* 2003;46:115–22.
- 20 Sethi S, D'Agati VD, Nast CC, et al. A proposal for standardized grading of chronic changes in native kidney biopsy specimens. *Kidney Int* 2017;91:787–9.
- 21 Shao D, Jimenez AL, Guerrero MS, et al. Factors Associated with Worsening Interstitial Fibrosis/Tubular Atrophy in Lupus Nephritis Patients Undergoing Repeat Kidney Biopsy. *Res Sq* 2024.:rs.3.rs-3867933.
- 22 He K, Gkioxari G, Dollár P, eds. In: Mask r-cnn. Proceedings of the IEEE international conference on computer vision; 2017
- 23 Zuiderveld KJ, ed. Editor Contrast Limited Adaptive Histogram Equalization. Graphics gems, 1994.
- 24 Tang ZY, Lin YC, Shen CC, eds. In: Dual-Path Convolutional Neural Network for Chronic Kidney Disease Classification in Ultrasound Echography. 2022 IEEE International Ultrasonics Symposium (IUS).
- 25 Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016:785–94.
- 26 Breiman L. Random forests. Mach Learn 2001;45:5–32.
- 27 Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 2017.
- 28 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
- 29 Moghazi S, Jones E, Schroepple J, et al. Correlation of renal histopathology with sonographic findings. *Kidney Int* 2005;67:1515–20.