frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 17 August 2022
DOI 10.3389/fphy.2022.920506

OPEN ACCESS

EDITED BY
Giovanni Mettivier,
Università degli Studi di Napoli Federico II, Italy

REVIEWED BY
Xiaopan Xu,
Fourth Military Medical University, China
Xun Sun,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Zhongwei Qiao,
zhongweiqiao@gmail.com
Jun Shi,
junshi@shu.edu.cn
Jia Rao,
jiarao@fudan.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

RECEIVED 14 April 2022
ACCEPTED 21 July 2022
PUBLISHED 17 August 2022

CITATION
Kang W, Ji M, Zhang H, Shi H, Xiang T, Li Y, Fang Y, Qi Q, Wang J, Shen J, Tang L, Liu X, Ye Y, Ge X, Wang X, Xu H, Qiao Z, Shi J and Rao J (2022), A novel clinical-radiomics model predicted renal lesions and deficiency in children on diffusion-weighted MRI. *Front. Phys.* 10:920506. doi: 10.3389/fphy.2022.920506

A novel clinical-radiomics model predicted renal lesions and deficiency in children on diffusion-weighted MRI

Weijie Kang[1,2,3†], Min Ji[4†], Huili Zhang[5†], Hua Shi[1,2,3], Tianchao Xiang[1,2,3], Yaqi Li[1,2,3], Ye Fang[1,2,3], Qi Qi[1,2,3], Junbo Wang[4], Jian Shen[6], Liangfeng Tang[6], Xiaoxiong Liu[7], Yingzi Ye[7], Xiaoling Ge[7], Xiang Wang[6], Hong Xu[1,2,3], Zhongwei Qiao[4*], Jun Shi[5*] and Jia Rao[1,3,8*]

[1]Department of Nephrology, Children's Hospital of Fudan University, Shanghai, China, [2]Shanghai Kidney Development and Pediatric Kidney Disease Research Center, Shanghai, China, [3]Shanghai Key Lab of Birth Defect, Children's Hospital of Fudan University, Shanghai, China, [4]Department of Radiology, Children's Hospital of Fudan University, Shanghai, China, [5]School of Communication and Information Engineering, Shanghai University, Shanghai, China, [6]Department of Urology, Children's Hospital of Fudan University, Shanghai, China, [7]Data Management Center, Children's Hospital of Fudan University, Shanghai, China, [8]State Key Laboratory of Medical Neurobiology, Institutes of Brain Science and School of Basic Medical Science, Fudan University, Shanghai, China

Background: Assessment of renal lesions and deficiency accurately remains critical in the diagnosis of congenital anomalies of the kidneys and urinary tracts (CAKUT) in children. Advanced imaging such as Magnetic resonance Imaging (MRI) and Diffusion weighted Imaging (DWI) allows structural and functional insufficiency to be detected. Currently, radiomics machine learning models are being explored as full-automated diagnostic tools. We aimed to develop a machine learning integrated radiomics model to predict renal anomalies and deficiency in children.

Methods: A retrospective study of 280 children with MRI/DWI were enrolled between 2018 and 202 at a children's hospital. A total of 1,037 radiomics features were extracted from the DWI images of each participant, which were divided into training set and test set (8:2 split). Using 5-fold cross-validated method, multiple machine learning algorithms were employed to predict renal lesions and deficiency when compared with the radiologist's diagnosis based on DWI, ^{99m}Tc-labeled dimercaptosuccinic acid (DMSA) SPECT cortical renal scintigraphy or ^{99m}Tc-labeled diethylenetriamine pentaacetate (DTPA) renal scan.

Results: For detecting the kidney lesions, the LASSO + Random Forest algorithm outperformed other classifiers with an accuracy of 0.750 (95% confidence interval, 0.734–0.766) and area under the curve (AUC) of 0.765 (95% confidence interval, 0.700–0.831). The performance of classifiers did not show a significant difference ($p > 0.05$) in detecting bilateral or unilateral kidney lesions by DWI scanning. The classifiers performed significantly better in bilateral kidney deficit than in unilateral kidney deficit ($p < 0.05$). We next built prediction models for renal deficiency using the radiomics signature

and clinical features compared to renal scintigraphy. The ensemble model had a high-test accuracy of 80.9% ± 4.2% and a sensitivity of 91.7% ± 7.1% with a moderate calibration.

Conclusion: An ensemble model integrated with DWI-radiomic and clinical features can be utilized to predict renal lesions and deficiency in children with CAKUT.

KEYWORDS

congenital anomalies of the kidneys and urinary tracts (CAKUT), machine learning, radiomics, diffusion weighted imaging (DWI), children

Introduction

Congenital anomalies of the kidneys and urinary tracts (CAKUT) account for roughly 20% of all birth defects [1]. CAKUT is not only associated with urinary tract infection (UTI), but also increases the risk of renal deficit in febrile UTI. CAKUT affects around 5 per 1,000 live births, accounting for 30%–50% chronic kidney disease (CKD) in children [2]. Therefore, early diagnosis of kidney anomalies and renal deficits is critical for CAKUT heath improvement.

Advanced imaging, including ultrasound, magnetic resonance imaging (MRI), and scintigraphy, allow for the retrieval of structural, functional, and molecular information that may be used to identify changes in renal tissue properties and functionality [3]. Diffusion weighted imaging (DWI) is a non-invasive magnetic resonance modality, which uses the motion of water molecules as a contrast to measure *in vivo* movement of water diffusion or Brownian motion. It reflects the microstructure of the tissue by describing the water restriction and diffusivity within this tissue. DWI has been described as a reliable biological marker of renal interstitial fibrosis, renal perfusion alterations or water handing [4]. Nevertheless, DWI quantification involves a lot of manual procedures, making it time consuming and observer dependent, limiting its clinical utility. Radiomics is a new tool for diagnosing renal disease that has been utilized successfully for renal cancer, diabetic nephropathy, polycystic kidney disease, and urolithiasis differentiation [5–8]. Radiomics is a method designed to retrieve a large number of quantitative features from medical images to aid diagnostic reasoning [9]. We hypothesized that DWI-based radiomics might be developed as a diagnostic assistance for CAKUT.

For early detection the renal deficiency before a loss of kidney function estimated with serological test, there is great interest in developing novel methods that can be used to detect renal deficiency in children at an early stage. Here we present a novel machine learning model for predicting renal anomalies and deficiency in children based on DWI radiomics and clinical features.

Materials and methods

Study population and clinical features

After receiving approval from the Research Ethics Board of Children's Hospital of Fudan University (NO. 2018-286), a total of 280 pediatric patients (age below 18 years old) were recruited in the study between January 2018 and December 2020. All participants and/or their legal guardians were fully informed about the study's aims and provided their written and/or verbal consent prior to the examination. The inclusion criteria were 1) age under 18 years old, 2) MRI including T1-weighted, T2-weighted and DWI were available, 3) the image quality was adequate for analysis, no motion or artifacts.

Basic clinical information included patients' age at examination, gender, and initial manifestation of urinary tract infection (UTI). Based on the diagnosis of DWI scanning from 3 independent senior radiologists, all the participants were divided into two groups: kidney lesions and non-lesion detected by MRI/DWI. Kidney function decline was defined as an abnormal value of split renal function using ^{99m}Tc-labeled dimercaptosuccinic acid (DMSA) SPECT cortical renal scintigraphy or ^{99m}Tc-labeled diethylenetriamine pentaacetate (DTPA) renal scan [10, 11]. Figure 1 showed the flowchart for the study.

Image processing

A 1.5-T MRI scanner (Avanto, Siemens Healthcare, Germany) was used. Signal reception was conducted with a six-element body matrix coil and 12-element design of the inbuilt spine matrix coil. During the scanning process, the individuals were required to lie flat on the examination bed, maintain free breathing, and keep their body motionless with sedation if necessary. The scan was performed by an experienced MR technician. Conventional coronal T2WI, axial T1WI, T2WI, and DWI sequences were performed. DWI applied a single-shot echo planar imaging (EPI) sequence with free-breathing. Chemical shift artifacts can be reduced by fat saturation technique. The images were acquired using a diffusion-

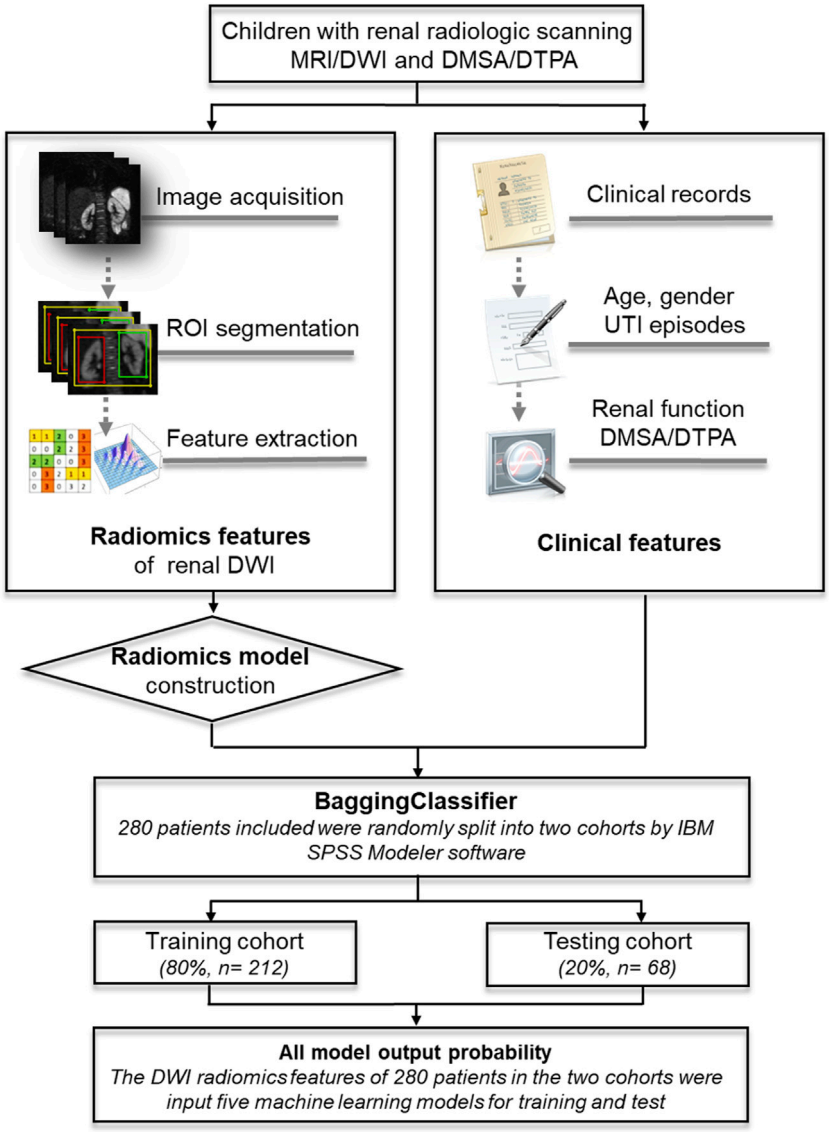

FIGURE 1
An illustration of the machine learning models based on DWI-radiomics in children. MRI, magnetic resonance imaging; DWI, Diffusion weighted imaging; DMSA, ^{99m}Tc-labeled dimercaptosuccinic acid SPECT cortical renal scintigraphy; DTPA, ^{99m}Tc-labeled diethylenetriamine pentaacetate renal scan.

weighted single shot echo-planar imaging fat saturated sequence. The repetition time (ms)/echo time (ms) was 3300/57; the flip angle was 90°; the results of two excitations were averaged; the field of view was 30 × 30 cm; the matrix size was 128 × 128; 10 coronal slices were acquired with a thickness of 4 mm; the parallel imaging (ASSET) factor was 3; and the acquisition time was 180 s. The b-values ranged from 0 to 700 (0, 10, 20, 50, 100, 180, 300, 420, 550, 700). Renal lesions identified by MRI/DWI include simple cysts, solid/mixed solid cysts, renal cortical defects, and local/diffuse signal intensity changes (high/low) in renal parenchymal.

All the images of MRI/DWI scanning were retrieved in DICOM format at their original dimensions and resolution. Two readers (a radiologist with 4 years of body MRI expertise and a medical student with MRI anatomy training) manually delineated the rectangular 2D ROI that included kidney slice by slice on a coronal DWI image, and then an experienced radiologist checked all the ROIs. The largest 2D ROI was then automatically selected by a self-developed software, and then used as the mask for all the slices that included kidney regions. After extracting the region in the mask of each slice, a cubic 3D ROI (or named VOI) was obtained by stacking all these slices

from a DWI image. The 3D ROI from each DWI image was then used for further radiomics analysis.

Radiomics feature extraction and feature selection

We employed an open-source python package called "PyRadiomics" to extract radiomic features from the DWI images, and all the results were collected in a form. The radiomics features were classified into 10 categories: First-order statistics (n = 18), shape (n = 14), texture derived from GLCM (n = 24), texture derived from GLSZM (n = 16), texture derived from GLRLM (n = 16), texture derived from GLDM (n = 14), texture derived from NGTDM (n = 5), wavelet-based features (n = 744), exponential-based features (n = 93), and logarithm-based features (n = 93) (Supplementary Table S1). Each feature was named by concatenating the image type from which the feature was extracted, feature group and feature name. For example, original_shape_Elong- ation was a feature extracted from the original image, shape group, and the feature name was Elongation.

In the feature selection, we first performed t-test analysis to pick features and those with p values < 0.05 were selected. The two independent sample t-test is used to test whether the mean and distribution of the two independent samples are significantly different. Several features involved averages in the two sets of samples of normal and abnormal patients, and the t-test was used to determine whether these features are substantially different. If there are significant differences, they can be reserved for classification tasks and this feature will be removed if there is no significant difference. T-test analysis was followed by the Lasso feature selection method, aiming to remove irrelevant features before classification. Lasso regression is based on ordinary least squares with L1 regular expression, the L1 regular expression is used to prevent the model from overfitting. Lasso can transform the value of irrelevant features into 0, thereby performing feature reduction and selecting important features for classification tasks.

Radiomics model construction and performance evaluation

To predict kidney lesions or deficiency, DWI-radiomics features were fed into four distinct models. We utilized the machine learning classifiers including Logistic Regression (LR), AdaBoost (AB), Support Vector Machines (SVM), and Random Forest (RF) from the same splits of patients to train and test on DWI features. All these classifiers were imported from a Python (version 3.6.5) machine learning library named scikit-learn (version 20.3). Models were trained with a batch size of 16. Early stopping was used with the patience parameter set to 50; finally, each model was allowed a maximum of 500 epochs of training if it was never stopped early. After 100 training trials, the model with the best validation accuracy was selected. During training, a predetermined probability of 0.5 was assigned to the final sigmoid activation neuron as a threshold for the classification of kidney lesion. The models were trained to maximize the accuracy of the model prediction. We split the training data set into five equally-sized parts, each exhibiting the same class distribution, and then the five-fold cross-validation was used to tune hyper-parameters.

The performance of the radiomics classifiers was firstly compared with that by radiologist diagnosis based on the DWI imaging data. It was subsequently compared with the diagnosis of kidney function decline according to the split renal function (SRF) estimated by DMSA or DTPA. To assess the performance of the classification, the accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), and the area under the curve (AUC) of its ROC curve with 95% confidence interval (95% CI) were calculated as evaluation (Supplementary Table S2). The classification results were expressed by the format of mean ± standard deviation (SD) over repeated runs. The confidence interval was calculated using the adjusted Wald method. p-values were calculated using a binomial test. A p-value of 0.05 was considered as the threshold for significance.

Development, performance of a clinical–radiomics model

To predict renal deficiency, a bagging classifier combined the clinical variable with radiomics signature to build an ensemble model. SPSS modeler (version 8.0, IBM) software randomly selected 80% of the dataset as the training group and 20% of the data as the test group to verify the trained model. Firstly, the five classifiers, including logistic regression (LR), decision tree (C5.0), SVM, neural network (Nnet), and RF were unutilized separately. To fit the weights of the models with 5-fold cross validation, we split the data into 5 subsets of roughly equal size and interactively used 4 subsets for interfold training and the fifth subset for interfold testing. We set the parameters as follows to enhance the performance of each classifier: Boosting was used in the C5.0 algorithm to improve model accuracy; binomial logistic regression with backward selection was used in the LR model; the polynomial function kernel was used as the kernel of the SVM; multilayer perceptron (MLP) networks were employed in Nnet; the MLP undersampling balanced training set was aggregated using the bagging ensemble method; 100 random tree number in RF model with the max feature included all the features input; the multi-model approach of ensemble is realized by combining

TABLE 1 Clinical characteristics of the patients.

Characteristic	Training cohort	Test cohort
case number	212	68
Age (years)		
Median (interquartile range)	1.6 (0.7, 5.8)	1.5 (0.6, 4.9)
Gender		
Male	121	32
Female	91	36
Clinical diagnosis		
CAKUT	97	28
Urinary tract infection	139	47
Tubulointerstitial lesions	2	1
Neuroblastoma	5	4
Radiological diagnosis		
Rena deficiency identified by DTPA or DMSA	140 (66.0%)	40 (66.0%)
Rena abnormalities identified by DWI	114 (53.8%)	33 (53.8%)

Abbreviation: CAKUT, congenital anomalies of the kidneys and urinary tracts; DMSA, ^{99m}Tc-labeled dimercaptosuccinic acid; DTPA, ^{99m}Tc-labeled diethylenetriamine pentaacetate; DWI, diffusion weighted imaging.

the prior models employed. Predictive performance was assessed using the AUC with 95%CI of its receiver operating characteristic (ROC AUC) curve and precision recall curve (PR AUC), SEN, SPE, PPV, NPV, and ACC. We compared the concordance between the renal deficiency as assessed by the best classifier in the five machine learning algorithms and radiological diagnosis of SRF. The DeLong method was used to compare the AUCs of the machine learning classifiers. A power calculation was performed to ensure that the test data set was sufficient to assess the AUC estimated from the training group.

Statistical analysis

All the statistical analysis was performed using SPSS (version 25.0, IBM, Armonk, New York) and SPSS Modeler (version 18.0). Graph was created using SPSS Modeler and GraphPad prism 8.3.0. To compare the differences in count data, Pearson's chi-squared test or Fisher's exact test were used. Statistical significance was defined as a p-value of less than 0.05. In the training set and testing set, the performance of the machine learning models was assessed with respect to their discrimination and calibration. The ROC AUC was calculated to assess the discrimination of the models, while the calibration was evaluated with the Hosmer-Lemeshow test. Cohen's kappa value was used to analyze the concordance between the best classifier and the radiologist's assessment of renal anomalies or deficiency. PASS v18 was used for sample size estimation (2018, NCSS, LLC. Kaysville, Utah, United States. www.ncss.com) and "Tests for One ROC Curve" function was used to perform the power calculation.

Results

Participants clinical characteristics

A total of 280 participants were enrolled in this study. There were 125 cases diagnosed with CAKUT initially, 140 cases diagnosed with UTI without CAKUT, 9 with neuroblastoma and 3 cases with tubulointerstitial lesions. Among the 186 participants with UTI history, 104 cases with only one episode of UTI, 32 cases with two episodes of UTI, and 50 cases with more than three episodes of UTI prior to the MRI/DWI scanning. The characteristics of the training and testing cohort were summarized in Table 1. There was no statistically significant difference in the clinical features between the two cohorts (Pearson test, $p > 0.05$).

Radiomics signature building and performance evaluation

Totally, 1,037 radiomics features were retrieved from each participant's DWI images. The Lasso feature selection approach was performed to identify the top ten radiomics features. The unsupervised clustering of these features was shown as a heat map (Figure 2). Univariate analysis revealed the top ten features in Supplementary Table S2. We selected the ten key features in the training set based on the regression algorithm in the fivefold cross-validation studies.

Based on the radiomics feature selection methods, the performance of the radiomics classifiers compared with the expert evaluation on DWI were presented in Table 2. With a ROC AUC of 0.765 (95% CI, 0.700–0.831) and ACC of 0.750 (95%CI, 0.734–0.766), the selection method LASSO + RF algorithm

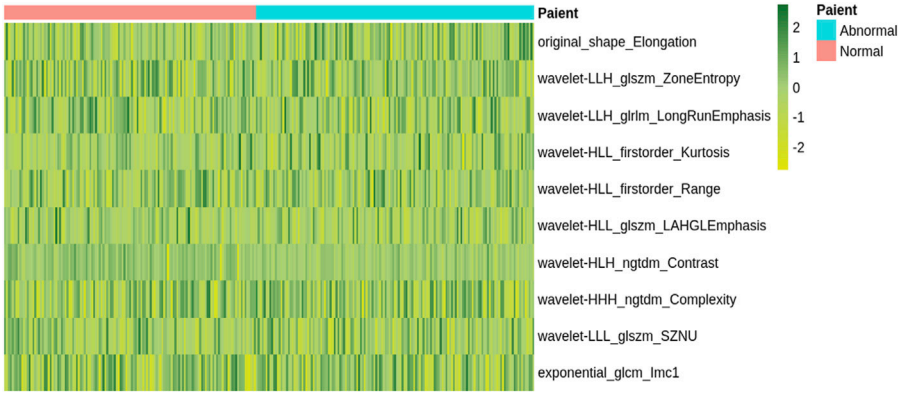

FIGURE 2
A heatmap shows the radiomic features. Each column and row correspond to one patient and z-score normalized radiomic feature, respectively.

TABLE 2 Parameters of detecting kidney lesions by radiomics analysis of DWI scanning.

Method	ACC (%)	SEN ± SD (%)	SPE ± SD (%)	PPV ± SD (%)	NPV ± SD (%)	AUC (95%CI)
LR	66.79 ± 2.04	70.70 ± 5.46	62.37 ± 2.94	67.48 ± 1.68	66.03 ± 3.17	0.697 (0.671, 0.723)
AB	72.14 ± 0.98	72.77 ± 0.80	71.38 ± 2.64	73.80 ± 1.31	70.32 ± 1.40	0.718 (0.666, 0.771)
SVM	73.57 ± 2.33	77.78 ± 3.58	68.74 ± 2.37	73.83 ± 3.01	73.26 ± 2.03	0.733 (0.697, 0.769)
RF	75.00 ± 1.26	76.82 ± 3.24	72.92 ± 3.11	75.86 ± 2.12	74.11 ± 1.44	0.765 (0.700, 0.831)
Detecting bilateral kidney lesions						
LR	65.25 ± 3.49	70.66 ± 11.14	69.06 ± 2.92	63.18 ± 5.88	75.52 ± 8.88	0.697 (0.616, 0.777)
AB	71.59 ± 3.84	73.90 ± 11.35	73.14 ± 11.86	68.42 ± 6.02	78.1 ± 10.12	0.736 (0.659,0.813)
SVM	70.16 ± 2.67	73.00 ± 4.96	74.56 ± 3.31	68.26 ± 7.49	78.48 ± 3.32	0.736 (0.659, 0.813)
RF	73.62 ± 2.79	77.20 ± 12.88	72.86 ± 3.68	68.34 ± 2.18	80.52 ± 11.31	0.744 (0.668, 0.821)
Detecting unilateral kidney lesions						
LR	66.23 ± 4.55	71.72 ± 12.22	62.40 ± 2.94	51.54 ± 5.95	80.00 ± 7.71	0.667 (0.591, 0.743)
AB	72.94 ± 4.23	73.86 ± 9.51	71.40 ± 2.65	59.14 ± 3.73	82.98 ± 6.60	0.726 (0.653, 0.798)
SVM	74.59 ± 3.00	82.50 ± 4.75	69.18 ± 3.14	60.06 ± 6.56	87.68 ± 2.24	0.760 (0.693, 0.828)
RF	75.89 ± 3.21	79.34 ± 7.68	72.94 ± 3.11	62.32 ± 2.57	86.16%5.96	0.759 (0.690, 0.828)

Abbreviation: AB, AdaBoost; ACC, the accuracy; AUC, area under the curve; DMSA, ^{99m}Tc-labeled dimercaptosuccinic acid; DTPA, ^{99m}Tc-labeled diethylenetriamine pentaacetate; DWI, Diffusion weighted imaging; LR, logistic regression; NPV, negative predictive value; PPV, positive predictive value; RF, random forest; SVM, support vector machines; SD, standard deviation; SEN, sensitivity; SPE, specificity.

outperformed other classifiers for identifying the kidney lesions. We examined the classifiers' predictive performance in patients with bilateral or unilateral kidney lesions (Figure 3). None of the classifiers showed a significant better performance in detecting the bilateral kidney lesions compared with unilateral lesions ($p > 0.05$, respectively, Table 2; Figure 3).

Furthermore, we tested the efficacy of the classifier for predicting the renal deficiency as identified by DTPA or DMSA. Similarly, the selection method LASSO + classifier RF presented the highest ROC AUC of 0.749 (95% CI, 0.690, 0.808) (Table 3). When we compared the models' performance in groups of bilateral and unilateral renal deficiency, we found that the LASSO + classifier RF in the bilateral renal deficiency group outperformed the unilateral renal deficiency group in terms of sensitivity, accuracy and AUC ($p < 0.05$, respectively, Table 3; Figure 3).

Development and performance of prediction models for renal deficiency

Firstly, a multivariate regression analysis was conducted integrating the radiomics and clinical features such as age,

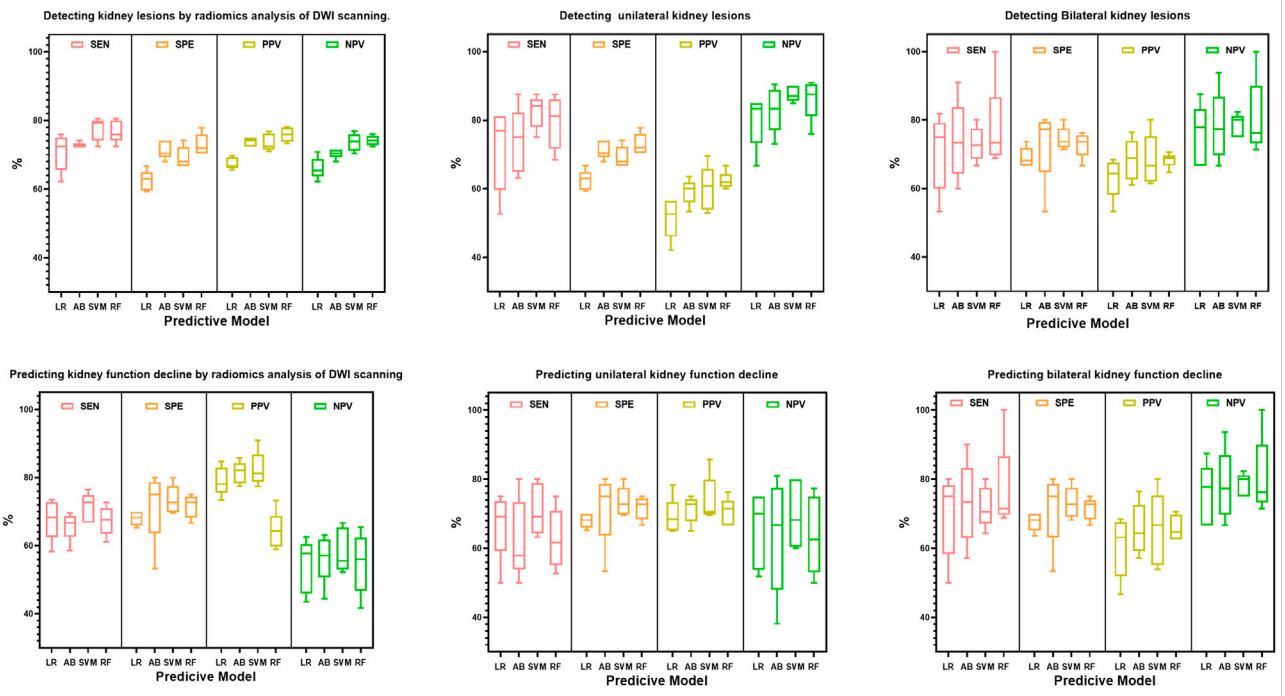

FIGURE 3
Box plot distributions of sensitivity (SEN), specificity (SPE), positive predictive value (PPV) and negative predictive value (NPV) under radiomics models for the prediction of bilateral or unilateral kidney lesions. The four machine learning classifiers including Logistic Regression (LR), AdaBoost (AB), Support Vector Machines (SVM) and Random Forest (RF) from the same splits of patients to train and test on DWI features.

TABLE 3 Parameters of detecting renal function decline identified by DTPA or DMSA.

Method	ACC ± SD (%)	SEN ± SD (%)	SPE ± SD (%)	PPV ± SD (%)	NPV ± SD (%)	AUC (95%CI)
LR	67.87 ± 3.79	67.8 ± 5.99	68.02 ± 2.10	67.48 ± 1.68	54.06 ± 7.86	0.666 (0.602, 0.730)
AB	68.21 ± 6.23	65.84 ± 4.28	71.90 ± 10.66	73.80 ± 1.31	54.46 ± 7.19	0.721 (0.660, 0.782)
SVM	71.79 ± 3.19	71.16 ± 4.33	73.46 ± 4.26	73.83 ± 3.01	58.46 ± 6.47	0.734 (0.673, 0.794)
RF	68.93 ± 2.99	67.34 ± 4.30	71.66 ± 3.34	75.86 ± 2.12	54.82 ± 8.88	0.749 (0.690, 0.808)
Detecting bilateral renal function decline						
LR	68.40 ± 6.26	69.63 ± 12.01	67.70 ± 2.66	60.41 ± 8.81	75.51 ± 8.90	0.686 (0.604, 0.767)
AB	72.41 ± 7.25	73.14 ± 11.90	71.67 ± 10.60	65.55 ± 7.44	78.00 ± 10.11	0.725 (0.646, 0.803)
SVM	72.66 ± 3.4.95	71.97 ± 5.88	73.18 ± 4.61	65.47 ± 10.71	78.47 ± 3.31	0.725 (0.646, 0.803)
RF	73.25 ± 3.49	76.82 ± 13.06	71.42 ± 3.20	65.81 ± 3.69	80.52 ± 11.31	0.734 (0.656, 0.811)
Detecting unilateral renal function decline						
LR	67.2 ± 4.80	66.97 ± 9.84	68.01 ± 2.09	69.06 ± 5.40	65.48 ± 11.02	0.670 (0.596, 0.744)
AB	66.98 ± 9.78	62.45 ± 11.45	71.9 ± 10.64	71.33 ± 3.87	63.46 ± 16.60	0.672 (0.599, 0.746)
SVM	71.84 ± 3.35	71.11 ± 7.47	73.46 ± 4.27	73.96 ± 6.79	69.81 ± 9.83	0.718 (0.648, 0.789)
RF	67.00 ± 5.02	62.71 ± 8.59	71.66 ± 3.36	70.48 ± 3.98	63.70 ± 11.34	0.672 (0.598, 0.746)

Abbreviation: AB, AdaBoost; ACC, the accuracy; AUC, area under the curve; DMSA, ^{99m}Tc-labeled dimercaptosuccinic acid; DTPA, ^{99m}Tc-labeled diethylenetriamine pentaacetate; DWI, diffusion weighted imaging; LR, logistic regression; NPV, negative predictive value; PPV, positive predictive value; RF, Random Forest; SVM, support vector machines; SD, standard deviation; SEN, sensitivity; SPE, specificity.

TABLE 4 Results from 5-fold cross-validation of predictions renal deficiency.

Training modality	ACC ± SD (%)	ROC AUC (95%CI)	PR AUC (95%CI)	SEN ± SD (%)	SPE ± SD (%)	PPV ± SD (%)	NPV ± SD (%)	Cohen's kappa value	*p*-value by Hosmer-Lemeshow test
LR	72.17 ± 7.82	0.664 (0.895, 0.965)	0.753 (0.674, 0.817)	71.43 ± 11.09	26.67 ± 14.12	64.71 ± 9.78	33.33 ± 21.36	0.112	0.387
C5	84.91 ± 4.55	0.855 (0.801, 0.900)	0.903 (0.842, 0.942)	68.00 ± 9.81	53.85 ± 18.58	70 ± 5.99	45 ± 9.16	0.185	0.594
Nnet	92.92 ± 13.21	0.694 (0.628, 0.756)	0.781 (0.705,0.842)	64.52 ± 15.52	38.89 ± 13.52	66.67 ± 17.8	35.29 ± 9.21	0.001	0.136
RF	62.50 ± 2.27	0.926 (0.882, 0.957)	0.949 (0.898–0.975)	68.97 ± 4.33	50 ± 8.65	68.97 ± 5.32	45.45 ± 6.74	0.184	0.785
SVM	87.74 ± 11.04	0.863 (0.810, 0.907)	0.894 (0.831, 0.935)	73.91 ± 14.35	45.45 ± 8.85	66.67 ± 7.43	40 ± 12.4	0.039	0.723
Ensemble	93.40 ± 2.62	0.937 (0.895, 0.965)	0.958 (0.908, 0.981)	92.31 ± 2.83	80 ± 11.79	89.19 ± 6.86	85.71 ± 9.12	0.751	0
Training modality	**ACC ± SD (%)**	**ROC AUC (95%CI)**	**PR AUC (95%CI)**	**SEN ± SD (%)**	**SPE ± SD (%)**	**PPV ± SD (%)**	**NPV ± SD (%)**	**Cohen's kappa value**	***p*-value by Hosmer-Lemeshow test**
LR	77.94 ± 16.7	0.748 (0.628, 0.846)	0.748 (0.593–0.859)	80.00 ± 15.19	40 ± 31.42	70 ± 20.68	50 ± 18.23	0.223	0.547
C5	79.41 ± 13.3	0.788 (0.671, 0.877)	0.801 (0.649–0.897)	66.67 ± 12.74	60 ± 20.33	75 ± 17.9	57.14 ± 24.99	0.131	0.581
Nnet	66.18 ± 14.26	0.643 (0.517, 0.755)	0.682 (0.524–0.807)	66.67 ± 14.51	42.22 ± 24.97	66.67 ± 18.97	40.00 ± 36.36	0.139	0.493
RF	55.88 ± 17.27	0.539 (0.414, 0.661)	0.613 (0.456–0.749)	66.67 ± 15.99	40 ± 32.67	62.5 ± 21.07	37.50 ± 12.88	0.052	0.934
SVM	77.94 ± 18.36	0.879 (0.777, 0.945)	0.882 (0.741–0.951)	81.82 ± 15.71	50 ± 37.12	85.71 ± 16.36	33.33 ± 33.32	0.013	0.579
Ensemble	80.88 ± 4.2	0.784 (0.667, 0.875)	0.779 (0.625–0.881)	91.67 ± 7.11	65.15 ± 2.13	77.5 ± 7.25	93.75 ± 23.66	0.996	0.133

Abbreviation: ACC, the accuracy; AUC, area under the curve; C5, decision tree of C5.0; LR, logistic regression; Nnet, neural networks; NPV, negative predictive value; PPV, positive predictive value; PR AUC, area under precision-recall curve; RF, random forest; ROC AUC, area under receiver operating characteristic curve; SVM, support vector machines; SD, standard deviation; SEN, sensitivity; SPE, specificity; 95% CI, 95% confidence interval.

10.3389/fphy.2022.920506

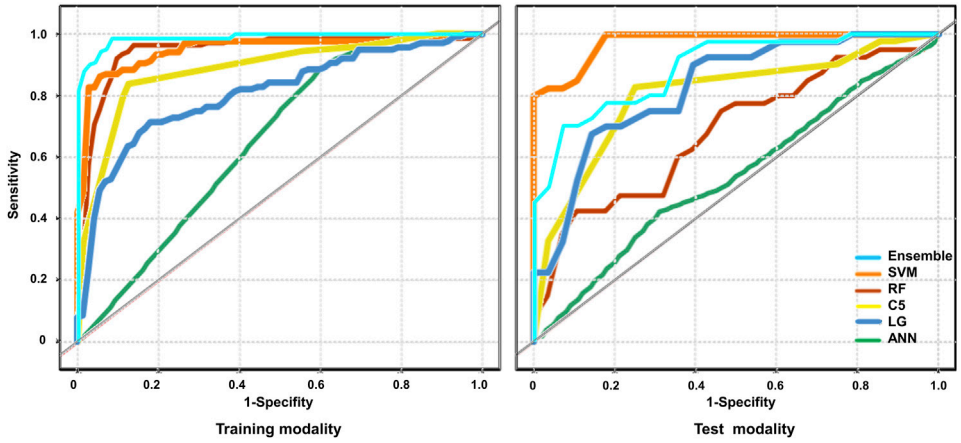

FIGURE 4
The performance of the models for the prediction of renal deficiency diagnosed by DMSA/DTPA. The Receiver operating characteristic (ROC) curves of the different models were presented across the training and test cohort. Logistic regression (LR), decision tree (C5.0), SVM, neural network (Nnet), Random Forest (RF) RF and ensemble models were unutilized separately.

gender and UTI episodes. The VIFs of the three potential predictors ranged from 1.07 to 1.63, indicating that multicollinearity was not present (Supplementary Figure S1). We also applied the five machine learning models to predict renal deficiency detected by DMSA or DTPA based on clinical features of gender, age, UTI episodes and selected DWI-radiomics features. The performance of the six models in the training and test sets are presented in Table 4 and Figure 4. Confusion matrices and reliability curves of all models in the training and test sets were displayed in Figure 5 and Supplementary Figure S2. The ROC AUCs in the trained model for LG, C5.0, Nnet, RF, and SVM were 0.664 (95%CI, 0.895–0.965), 0.855 (95%CI, 0.801–0.900), 0.694 (95%CI, 0.628–0.756), 0.926 (95%CI, 0.882–0.957), 0.863 (95%CI, 0.810–0.907), respectively. Application of the five machine learning algorithms in the test cohort yield high AUCs of SVM and RF with 0.926 (95%CI, 0.882–0.957), 0.879 (95%CI, 0.777–0.945), respectively.

An ensemble model was built by combining the five prior machine learning models including LR, C5.0, RF, Nnet, and SVM. The performance was compared with the other independent machine learning algorithms in predicting renal deficiency (Table 4). The ensemble model had the advantage of a high-test accuracy (80.88% ± 4.2%) and sensitivity (91.67% ± 7.11%). In the training set, the ensemble model achieved a ROC AUC of 0.937 (95%CI, 0.895–0.965), while in the testing set, it had a ROC AUC of 0.784 (95%CI, 0.667–0.875). Strong concordance between the ensemble model and renal deficiency identified by DMSA/DTPA was presented with a Cohen's kappa value of 0.906. Good consistency was also observed between the ensemble prediction model and the renal deficiency in the test set with nonsignificant p values (0.133) derived from the Hosmer-Lemeshow test.

Based on the AUC of the ensemble model in the training set and testing set, we set the AUC0 = 0.937, the AUC1 = 0.784, α = 0.05, false positive rate limited: 0.01–0.20 for power calculation. The results revealed that when the target power was 0.80, the sample sizes of the test sets were 60 or 50, respectively. Therefore, the test sets were sufficient to evaluate the AUC calculated from the training data.

Discussion

In this retrospective study, we illustrated that radiomic features of DWI appear to be reliable imaging indicators of renal lesions. A machine leaning approach integrated DWI-radiomics features and clinical variables was applied to predict renal deficiency with good performance.

There have been a few studies employing radiomics models as computer assisted diagnostic tools for renal deficits using DWI scanning [3]. The detection of renal parenchymal lesions by MRI or DWI scanning in children with CAKUT is very time consuming. It requires experienced and mature radiologists to avoid misdiagnosis. Radiomics texture analysis can provide quantitative information that is normally invisible to the naked eyes of a radiologist. Texture analysis approaches describe inter-relationships between the pixels and the gray-level frequencies within an image that cannot be sensed visually to quantify the variability of the distribution of pixels and their spatial arrangement. Owing to fibrosis as the hallmark of renal damage and the heterogeneity of the renal parenchyma, texture analysis on medical images of kidneys may be a useful predictor of renal lesions or deficiencies. Our findings confirmed the feasibility and

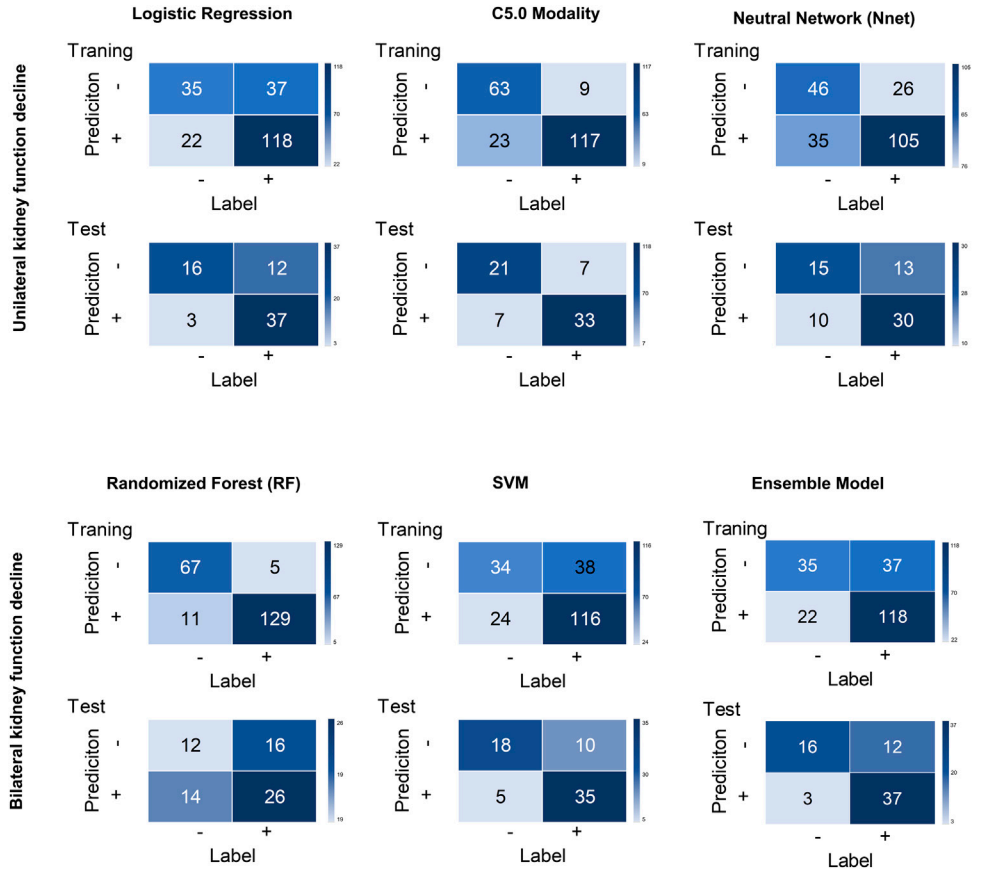

FIGURE 5
Confusion matrices for all models across the training and test cohort. Logistic regression (LR), decision tree (C5.0), SVM, neural network (Nnet), Random Forest (RF) RF and ensemble models were unutilized separately.

efficacy of radiomics classifiers to detect renal parenchymal lesions by DWI scanning. Instead of estimating radiomics with region-based segmentation of kidneys, we manually established the square areas in the bilateral kidneys and applied the radiomics over the entire region. For each scenario, manually establishing the ROI took only 5–10 s. After training and cross-validating the four common machine learning classifiers based on renal DWI images, we illustrated that the LASSO + RF algorithm performed best with a high AUC and ACC compared to expert evaluation on renal DWI in children. RF is a well-known machine learning algorithm for classification tasks that has an inherent resistance to overfitting. It builds multiple decision trees using random data points from the dataset and enhances the final prediction performance [12]. In this study, we applied stratified 5-fold cross-validation, which randomly split all the data into five parts and then held out 20% of the testing data, repeated five times. It might be difficult to identify the bilateral diffuse lesions in the kidneys without contralateral comparison by the naked eye. We found that the radiomics models had comparable accuracy and AUC for identification of kidney lesions by DWI when comparing the unilateral and bilateral lesions group. Hence, radiomics classifiers combined with machine learning techniques have been shown to be a useful complementary tool for clinicians, particularly in the early detection of renal lesions in childhood.

The split renal function can be used to measure the renal function through renal scintigraphy, such as DMSA and DTPA. DMSA and DTPA are regarded as the “gold standards” for diagnosing renal insufficiency in children due to their excellent diagnostic sensitivity [10, 11]. However, the procedure of DMSA or DTPA is invasive and involves radiation exposure. MRI/DWI outperformed DMSA/DTPA in terms of safety without the need for contrast agent injection. Calculating GFR with dynamic contrast-enhanced MRI in children, especially infants, is more challenging than in adults because of the faster inherent respiratory and cardiac motion [13]. It is essential to improve the accuracy of MRI/DWI in the diagnosis of renal deficiency. In the current study,

the LASSO + classifier RF or SVM models not only had the prediction accuracy of 68.9%–71.8% for renal deficiency identified by DTPA or DMSA, but they also performed better in bilateral kidney deficiency than in unilateral kidney deficiency. Even when the split renal function shows normal variation of 50%/50%–44%/56% (one kidney compared with the other), radiomics models may help to overcome the misdiagnosis of bilateral renal deficiency. Further study at 3-T field strength MRI in children will provide more high-quality radiological information to conduct quantitative analysis of renal function.

To further improve the diagnostic accuracy and sensitivity for predicting renal deficiency, we applied models combined the radiomics features with clinical variables when compared to calculation of split renal function by DMSA or DTPA. Ideally, a deep learning model would be used as a risk stratification tool to assist clinicians and patients in detecting early changes in renal function. To this purpose, the most essential statistic is sensitivity, because higher sensitivity, means that more children will receive an early diagnosis of renal damage or insufficiency. In addition to the radiomics signature, we involved the clinical variables such as age, gender and UTI episodes in machine learning models. This is in accordance with the viewpoint that UTI in childhood is associated with long-term outcome of renal function decline [14].

Compared with previous machine-learning studies, our work has several differences. First, we chose DWI that offers more information on renal parenchymal changes than simple attenuation differences measured in Hounsfield units on CT. Our machine learning algorithms retrieved radiomic features and improved the accuracy for clinical utilization. Second, to the best of our knowledge, this is the first study focusing on the role of DWI-based radiomics in children with CAKUT [15, 16]. All available normal or abnormal findings in kidneys from children who underwent MRI/DWI and DMSA/DTPA were included in our cohort. Most of the previous studies, on the other hand, only covered oncology or a limited number of CKD cases. Third, we employed clinical-radiomics fused data-based modalities to compare the prediction accuracy to radiologist diagnosis based on DMSA or DTPA. An ensemble model combining multiple machine learning models was transformed into the most optimal model in renal deficiency prediction with a high-test accuracy (ACC, 80.9%) and sensitivity (91.7%) after 5-fold cross-validation. Our cross-validation findings achieved a test AUC of 0.78 and a test PR AUC of 0.78 on average across all folds, confirming the model-building process.

There were several limitations to this study. First, overfitting a machine learning model is easy, especially when training with small amounts of data. Although the test accuracy reported for the best model was 80.9%, generalization is questionable due to the small cohort size and the lack of external validation. Despite the fact that we included 280 cases in our cohort, which is more than any prior machine learning works on CAKUT or CKD [3], algorithm development might benefit from a larger patient cohort, especially given the variability of image capture settings among institutions. Second, manual set for ROI was performed by students and radiologists reading the DWI images. On CT or MRI, automatic segmentation based on deep learning has been successfully established with equivalent accuracy to expert segmentation in kidneys [5, 17, 18]. Due to the small volume in children's kidneys, it is difficult to tell the kidneys apart from other organs in the body which may have comparable intensity, especially in CAKUT. We use the rectangle ROI to locate the bilateral kidneys which takes 5–10 s which can be substituted by automatic kidney segmentation in future. And time saving analysis should be performed when deployed into the clinical routine. In our initial exploration of the DWI-radiomics, the ROI selection is based on experience. In the future work, we can compare different choices of ROI, including manual sketching or automatic segmentation of the planar ROI, twelve-layer concentric objects (TLCO) [19], and three-dimensional volumes-of-interest (VOIs) [20], to see the difference in classification results. It may also provide novel sights on perinephric changes. Third, as a single-center retrospective study, it might result in a potential selection bias. The diagnostic performance of DWI-based radiomic model needs further study in multi-centers. Fourth, this study is based on DWI-derived radiomics combined with clinical features to predict renal deficiency. Multiparametric MRI-derived radiomics with united model based on T1WI, T2WI and ADC sequences may provide more successful prediction model for CAKUT. Future study on multiparametric MRI-derived radiomics will be conducted.

In conclusion, we established a machine learning model using the DWI-radiomics and clinical features to predict renal lesions and deficiency with high accuracy and sensitivity comparable with DMSA/DTPA-based diagnosis. If further validated, the algorithm can contribute to the early detection of renal lesions and predict renal deficiency in CAKUT.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by NO. 2018_28. Written informed consent to

participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

JR and JS led and supervised the project and were involved in all aspects of the study. JR, JS, and ZQ conceived and designed the study. WK, HZ, MJ, HS, TX, YL, YF, QQ, JW, JS, LT, XL, YY, XG, and XW acquired the data. WK and HZ processed the images. JR and HZ analyzed and interpreted the data. All authors contributed to the article and approved the final version.

Funding

JR is supported by a grant from the National Key Research and Development Program of China (2021YFC2701100), a grant from Clinical Research Plan of SHDC (SHDC2020CR2064B).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2022.920506/full#supplementary-material

References

1. Stonebrook E, Hoff M, Spencer JD. Congenital anomalies of the kidney and urinary tract: A clinical review. *Curr Treat Options Pediatr* (2019) 5(3):223–35. doi:10.1007/s40746-019-00166-3

2. Murugapoopathy V, Gupta IR. A primer on congenital anomalies of the kidneys and urinary tracts (CAKUT). *Clin J Am Soc Nephrol* (2020) 15(5):723–31. doi:10.2215/CJN.12581019

3. Alnazer I, Bourdon P, Urruty T, Falou O, Khalil M, Shahin A, et al. Recent advances in medical image processing for the evaluation of chronic kidney disease. *Med Image Anal* (2021) 69:101960. doi:10.1016/j.media.2021.101960

4. Sulkowska K, Palczewski P, Duda-Zysk A, Szeszkowski W, Wojcik D, Kownacka-Piotrowska D, et al. Diffusion-weighted MRI of kidneys in healthy volunteers and living kidney donors. *Clin Radiol* (2015) 70(10):1122–7. doi:10.1016/j.crad.2015.05.016

5. Abdeltawab H, Shehata M, Shalaby A, Khalifa F, Mahmoud A, El-Ghar MA, et al. A novel CNN-based CAD system for early assessment of transplanted kidney dysfunction. *Sci Rep* (2019) 9(1):5948. doi:10.1038/s41598-019-42431-3

6. CongHua LQ-Q, Huang Z-Q, Ma Q-L, Wang X-M, Huang C-C, Xu J-X, et al. A radiomics method based on MR FS-T2WI sequence for diagnosing of autosomal dominant polycystic kidney disease progression. *Eur Rev Med Pharmacol Sci* (2021) 25(18):5769–80. doi:10.26355/eurrev_202109_26795

7. Pei X, Wang P, Ren J-L, Yin XP, Ma LY, Wang Y, et al. Comparison of different machine models based on contrast-enhanced computed tomography radiomic features to differentiate high from low grade clear cell renal cell carcinomas. *Front Oncol* (2021) 11:659969. doi:10.3389/fonc.2021.659969

8. Zheng J, Yu H, Batur J, Shi Z, Tuerxun A, Abulajiang A, et al. A multicenter study to develop a non-invasive radiomic model to identify urinary infection stone *in vivo* using machine-learning. *Kidney Int* (2021) 100(4):870–80. doi:10.1016/j.kint.2021.05.031

9. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: Bridging imaging and genomics. *Abdom Radiol (Ny)* (2019) 44(6):1960–84. doi:10.1007/s00261-019-02028-w

10. Mandell GA, Eggli DF, Gilday DL, Heyman S, Leonard JC, Miller JH, et al. Society of nuclear medicine procedure guideline for renal cortical scintigraphy in children In: *Society of Nuclear Medicine Procedure Guidelines Manual August*. Available at: https://humanhealth.iaea.org/HHW/NuclearMedicine/PaediatricsandNephrourology/GuidelinesandLiterature/SNMGuidelinesPaediatrics/Renal_Cortical_Scintigraphy.pdf (2003). 195–198.

11. Treves ST. *Pediatric nuclear medicine and molecular imaging*. New York, NY: Springer New York (2014).

12. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer* (2018) 18(8):500–10. doi:10.1038/s41568-018-0016-5

13. Kurugol S, Afacan O, Lee RS, Seager CM, Ferguson MA, Stein DR, et al. Prospective pediatric study comparing glomerular filtration rate estimates based on motion-robust dynamic contrast-enhanced magnetic resonance imaging and serum creatinine (eGFR) to 99mTc DTPA. *Pediatr Radiol* (2020) 50(5):698–705. doi:10.1007/s00247-020-04617-0

14. Calderon-Margalit R, Golan E, Twig G, Leiba A, Tzur D, Afek A, et al. History of childhood kidney disease and risk of adult end-stage renal disease. *N Engl J Med* (2018) 378(5):428–38. doi:10.1056/nejmoa1700993

15. Chen C-J, Pai T-W, Hsu H-H, Lee CH, Chen KS, Chen YC. Prediction of chronic kidney disease stages by renal ultrasound imaging. *Enterprise Inf Syst* (2020) 14(2):178–95. doi:10.1080/17517575.2019.1597386

16. Ding J, Xing Z, Jiang Z, Zhou H, Di J, Chen J, et al. Evaluation of renal dysfunction using texture analysis based on DWI, BOLD, and susceptibility-weighted imaging. *Eur Radiol* (2019) 29(5):2293–301. doi:10.1007/s00330-018-5911-3

17. Homayounieh F, Doda Khera R, Bizzo BC, Ebrahimian S, Primak A, Schmidt B, et al. Prediction of burden and management of renal calculi from whole kidney radiomics: A multicenter study. *Abdom Radiol (Ny)* (2021) 46(5):2097–106. doi:10.1007/s00261-020-02865-0

18. Boer Ade, Harteveld AA, Stemkens B, Blankestijn PJ, Bos C, Franklin SL, et al. Multiparametric renal MRI: An intrasubject test-retest repeatability study. *J Magn Reson Imaging* (2021) 53(3):859–73. doi:10.1002/jmri.27167

19. Li L-P, Milani B, Pruijm M, Kohn O, Sprague S, Hack B, et al. Renal BOLD MRI in patients with chronic kidney disease: Comparison of the semi-automated twelve layer concentric objects (TLCO) and manual ROI methods. *Magn Reson Mater Phy* (2020) 33(1):113–20. doi:10.1007/s10334-019-00808-5

20. Vicente EM, Lodge MA, Rowe SP, Wahl RL, Frey EC. Simplifying volumes-of-interest (VOIs) definition in quantitative SPECT: Beyond manual definition of 3D whole-organ VOIs. *Med Phys* (2017) 44(5):1707–17. doi:10.1002/mp.12164