MAJOR ARTICLE

Radiomics and Clinical Features for Distinguishing Kidney Stone–Associated Urinary Tract Infection: A Comprehensive Analysis of Machine Learning Classification

Jianjuan Lu,[1,a] Kun Zhu,[2,a] Ning Yang,[1] Qiang Chen,[1] Lingrui Liu,[1] Yanyan Liu,[1,3,4,5,6] Yi Yang,[1] and Jiabin Li[1,3,4,5,6]

[1]Department of Infectious Diseases, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [2]Department of Orthopedics, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [3]Anhui Province Key Laboratory of Infectious Diseases, Anhui Medical University, Hefei, China, [4]Institute of Infectious Diseases, Anhui Medical University, Hefei, China, [5]Institute of Bacterial Resistance, Anhui Medical University, Hefei, China, and [6]Anhui Center for Surveillance of Bacterial Resistance, The First Affiliated Hospital of Anhui Medical University, Hefei, China

Background. This study investigated the abilities of radiomics and clinical feature models to distinguish kidney stone–associated urinary tract infections (KS-UTIs) using computed tomography.

Methods. A retrospective analysis was conducted on a single-center dataset comprising computed tomography (CT) scans and corresponding clinical information from 461 patients with kidney stones. Radiomics features were extracted from CT images and underwent dimensionality reduction and selection. Multiple machine learning (Three types of shallow learning and four types of deep learning) algorithms were employed to construct radiomics and clinical models in this study. Performance evaluation and optimal model selection were done using receiver operating characteristic (ROC) curve analysis and Delong test. Univariate and multivariate logistic regression analyzed clinical and radiomics features to identify significant variables and develop a clinical model. A combined model integrating radiomics and clinical features was established. Model performance was assessed by ROC curve analysis, clinical utility was evaluated through decision curve analysis, and the accuracy of the model was analyzed via calibration curve.

Results. Multilayer perceptron (MLP) showed higher classification accuracy than other classifiers (area under the curve (AUC) for radiomics model: train 0.96, test 0.94; AUC for clinical model: train 0.95, test 0.91. The combined radiomics-clinical model performed best (AUC for combined model: train 0.98, test 0.95). Decision curve and calibration curve analyses confirmed the model's clinical efficacy and calibration.

Conclusions. This study showed the effectiveness of combining radiomics and clinical features from CT scans to identify KS-UTIs. A combined model using MLP exhibited strong classification abilities.

Kidney stone disease, a key subtype of urolithiasis, poses a serious threat to human health worldwide [1]. The incidence of kidney stone disease is closely associated with numerous factors, such as genetics, sex, age, occupation, metabolic disorders, medication, diet, urinary tract obstruction, and infection [2]. However, kidney stone–associated urinary tract infections (KS-UTIs) represent a unique and challenging disorder, with a cumbersome diagnostic process and suboptimal treatment efficacy. Despite appropriate treatment, KS-UTI patients frequently experience poor recovery, with a recurrence rate approaching 50% over 5–10 years. Patients with severe disease may experience renal failure, loss of renal function, and death [3–6]. Therefore, the healthcare costs associated with KS-UTI are very high [7], representing a substantial threat to patient safety, which requires close attention.

As the prevalence of patients with KS-UTIs continues to rise, conventional examination methods often impede timely treatment due to their lengthy duration and cumbersome procedures. Consequently, these traditional approaches increasingly fail to meet the needs of patients. There is an urgent demand for a noninvasive and portable method to determine whether kidney stones in certain patients are accompanied by infection. Radiomics, an emerging noninvasive method, uses high-throughput mining of quantitative features from medical images to obtain biomarkers that can guide clinical management [8]. Radiomics can be regarded as a "digital biopsy" that can provide an extensive description of pathological phenotype and spatial heterogeneity for disease lesions in various clinical settings [9]. In recent years, radiomics has attracted increasing attention in the field of renal diseases. Previous studies have demonstrated the practical value of radiomics in urological diseases [10]. To our knowledge, there have been few reports concerning potential applications of radiomics in KS-UTI management, and there is a lack of relevant published research. Noncontrast computed tomography (CT), the gold standard for urolithiasis detection, can provide information regarding stone quantity, size, and location [11]. Moreover, different machine learning models, including shallow learning and deep learning, based on radiomic clinical biomarkers, which are widely used across various medical fields, are highly valuable. This analysis can aid in identifying the most significant or sensitive models for specific diseases [12]. It is worth noting that, despite the significant advantages of radiomics methods, it is essential to comprehensively consider multiple clinical indicators in practical applications. Clinical indicators, such as patients' age, sex, medical history, symptoms, and laboratory test results, are indispensable for accurate diagnosis. Integrating these indicators with radiomics data may provide a more comprehensive and profound understanding of the disease.

Our objectives encompass 3 primary facets. First, we aim to determine whether machine learning models can effectively leverage both radiomic features and clinical indicators to identify

patients with KS-UTIs. Second, we seek to elucidate the extent to which radiomic and clinical features provide insights into the pathogenesis of the disease. Last, we endeavor to assess whether the integration of radiomic and clinical features can enhance classification accuracy. To achieve these goals, it is imperative to rigorously validate relevant radiomics methodologies and conduct a comprehensive investigation into the potential associations between radiomic features and clinical characteristics within the context of spatiotemporal molecular imaging. This initiative holds significant promise for advancing clinical translation and deepening our understanding of the disease [13].

MATERIALS AND METHODS

Participants

This retrospective study included consecutive patients with urolithiasis who underwent pretreatment CT scans in our hospital from August 2021 to May 2023. Inclusion criteria were complete medical records and follow-up data, age 18–85 years, stone size ≥2 mm without any coagulation disorder, and no history of other malignant tumors. Exclusion criteria were poor image quality or the presence of artifacts; age <18 years;

pregnancy or renal anatomical anomalies; and/or solitary kidney, urinary diversion, or lower urinary tract reconstruction, which may potentially influence the diagnostic accuracy of the CT images (Figure 1).

Patient image and clinical data were obtained from our hospital's routine clinical records and Picture Archiving and Communication System. We retrospectively analyzed radiomics parameters and relevant clinical parameters for all participants, including age, sex, stone composition, white blood cell count, urine culture, urine pH, and serum creatinine. CT images were evaluated and approved by 2 experienced radiologists who were blinded to the clinical data of all participants. We categorize kidney stones into KS-UTIs and noninfected kidney stones, based on the positivity or negativity of the urine culture.

Image Acquisition

CT scans of the urinary tract in all participants were conducted via spiral CT with a matrix of 512×512, tube voltage of 120 kVp, automatic tube current, 300 mA; rotation time, 0.5 seconds, and 0.94:1 pitch; scan slice thickness, 5 mm, 24×1.2 mm detector collimation; and reconstruction thickness, 1.25 mm with a 1-mm overlap and soft tissue reconstruction
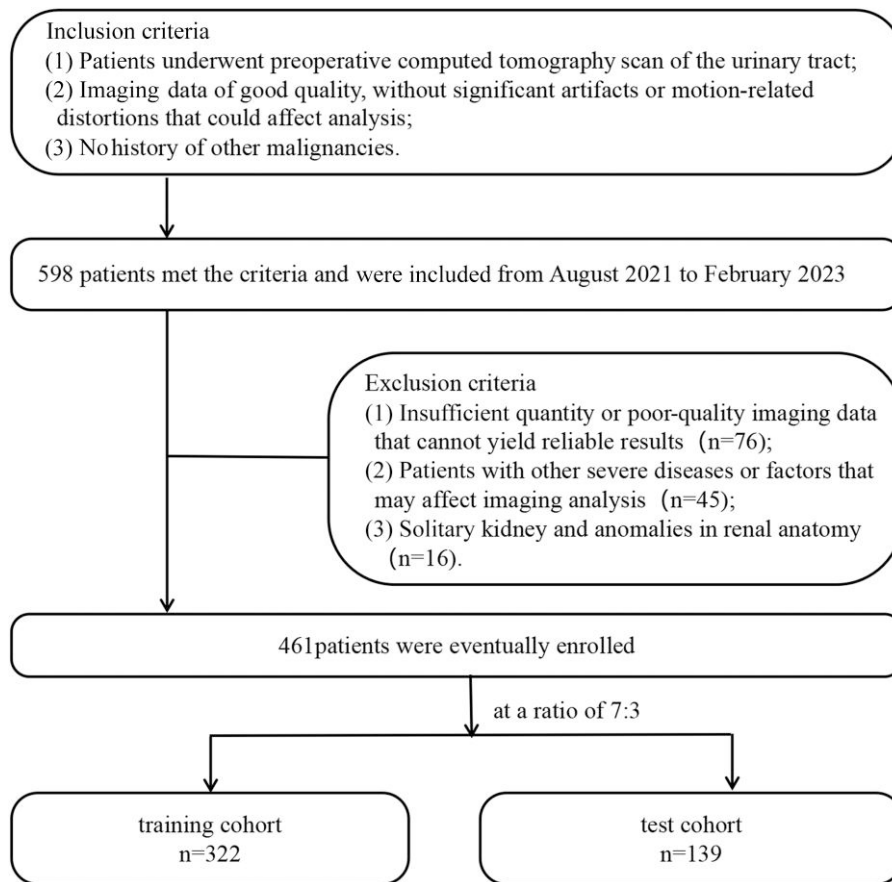

Figure 1. Flowchart for selecting the study population.

algorithm using vendor-specific iterative reconstruction techniques (ASIRv, GE Healthcare). The scanning range extended from the adrenal glands to the level of the pubic symphysis. Moreover, to mitigate potential biases and variations arising from human factors, all CT imaging scans in this study were conducted on a single, dedicated scanner, thereby ensuring consistency in both hardware and software configurations. The image acquisition process was standardized and rigorously implemented by a single, experienced technician trained to uphold uniform scanning protocols and techniques across all participants. This methodology not only minimized the risk of interoperator variability but also ensured that the quality and interpretability of the CT scans remained consistent throughout the duration of the study.

Image Segmentation and Radiomics Feature Extraction

The radiomics workflow was performed using an existing standardized procedure [14], as shown in Figure 2. All participants' images were stored in Digital Imaging and Communications in Medicine (DICOM) format and saved with standard soft tissue settings (window width, 400 Hounsfield units; window level, 40 Hounsfield units). Two radiologists (A and B) manually segmented the 3-dimensional region of interest (ROI) of stones in the urogenital system in a layer-by-layer manner using MRIcroGL (https://github.com/rordenlab/MRIcroGL). MRIcroGL, a free, open-source, and lightweight software, offers the capability to convert DICOM files imported from Picture Archiving and Communication System into Neuroimaging Informatics Technology Initiative (NifTI) format or directly import DICOM files. When delineating the ROI for lesions, MRIcroGL provides various colors, effectively facilitating the marking of distinct ROI and assisting in complex renderings; they ensured that adjacent normal tissues were not included in the segmentation. Radiomics features were extracted using the Radiomics package (https://pyradiomics.readthedocs.io/en/latest/features.html) in 3DSlicer software (https://www.slicer.org). Image and data preprocessing were performed by resampling to a voxel size of 1 mm^3 and z-score normalization, ensuring reliability and reproducibility of the subsequent analyses. Intra- and intergroup intraclass correlation assessments were performed for 436 features extracted from the 2 sets of ROIs. Features with intraclass correlation coefficient >0.7 were assumed to have good consistency and were subjected to further analysis. Finally, 414 radiomics features were obtained, including 14 shape- and morphology-based features, 18 first-order statistics features, 24 gray-level co-occurrence matrix features, 14 gray-level dependency matrix features, 16 gray-level run-length matrix features, 16 gray-level size region matrix features, and 5 neighbor gray-level tone difference matrix features. Aside from the original image, features were also extracted from Laplacian of Gaussian (LoG)–filtered images. For the LoG filter, images were filtered using a 3D LoG filter implemented in SimpleITK and by changing sigma values to 5.0, 4.0, and 3.0 mm, yielding another 3 derived images. Given the potential for overfitting with high-dimensional features, we opted not to extract wavelet features [15]. Instead, our primary focus is on evaluating the classification performance of KS-UTIs.

The radiomics features were normalized to the range of 0–1 using the z-score standardization method. Similar features
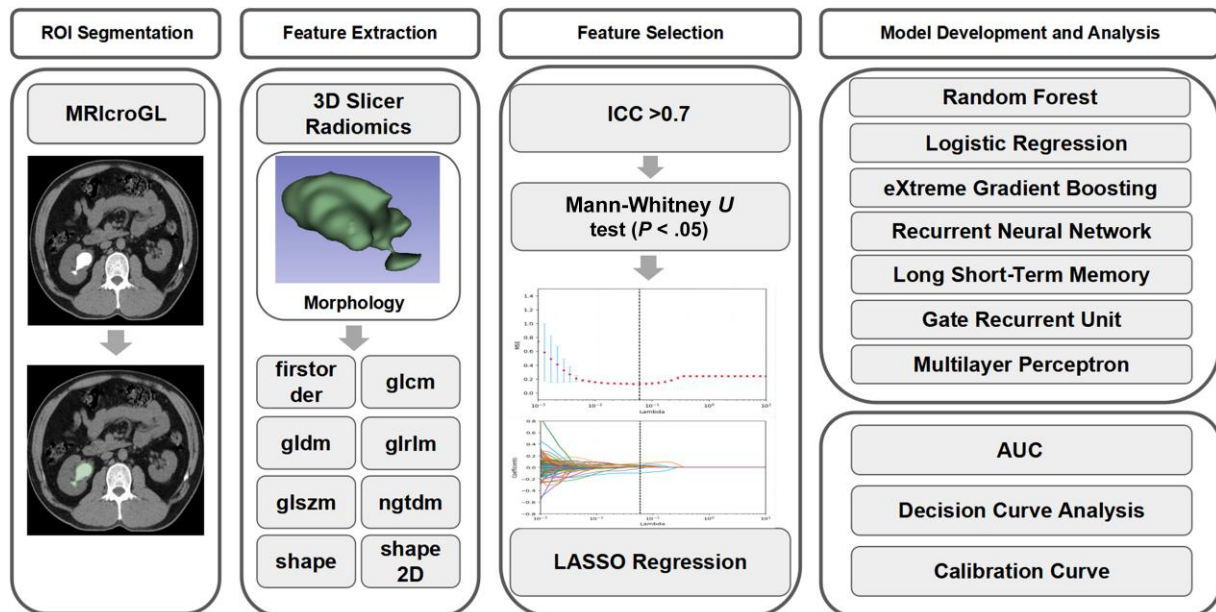

Figure 2. Workflow of this study. Abbreviations: AUC, area under the curve; ICC, intraclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; ROI, region of interest.

were removed by using the Mann-Whitney *U* test, which preserves distinct features by comparing the median differences of 2 independent samples. Finally, we used a 5-fold cross-validation least absolute shrinkage and selection operator (LASSO) regression model to select features with nonzero coefficients. By adding an L1 regularization term to the least squares algorithm, the LASSO algorithm can select the most representative radiomics features and use them for subsequent analysis.

Screening of Optimal Classifier and Construction of a Combined Model

In this study, we used the training set of participant data to identify key features and develop predictive algorithms, then used the test set of participant data to evaluate predictive performance. Based on the selected radiomics features, we employed 3 shallow learning methods and 4 deep learning methods to train a classifier model capable of accurately identifying infectious stones. The methods utilized include logistic regression (LR), random forest (RF), Xtreme gradient boosting (XGBoost), recurrent neural networks (RNN), long short-term memory (LSTM), gated recurrent unit (GRU), and multilayer perceptron (MLP). These models have been widely validated for their accuracy and robustness in classification tasks [16–20]. For shallow learning models, we performed appropriate parameter optimization. In the case of deep learning models, we uniformly adopted the RMSProp optimizer, which adaptively adjusts the learning rate for each parameter by considering the magnitude of recent gradients. This approach mitigates several issues arising from a global learning rate [21]. The learning rate was consistently set to 0.001. Furthermore, we employed the binary cross entropy loss function as our chosen loss function. The primary codes used in this study have been open-sourced and are available at http://github.com/zhukun990427/KS-UTIs.

We randomly sampled 70% of the data from all subjects to form the training set for 5-fold cross-validation, while the remaining 30% served as the test set. During the 5-fold cross-validation process, the training set data were randomly divided into 5 disjoint subsets of roughly equal size. In each of the 5 iterations, 1 subset was designated as the validation set, while the remaining 4 subsets were combined to form the training set, ensuring comprehensive model training. In this study, we compared the diagnostic performance of various models using several metrics: the area under the receiver operating characteristic (ROC) curve (AUC); accuracy, alongside the mean and standard deviation of accuracy across each cross-validation iteration (Acc_mean, Acc_std); and sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Additionally, we introduced 2 new metrics: the Dice similarity coefficient (DSC) and the Jaccard index [22]. The DSC, typically used to calculate the similarity between 2 samples, is computed as 2 * TP / (2 * TP + FP + FN), where TP represents true

Table 1. Analysis of Positive Urine Culture Results in Patients With Kidney Stones

Type of Bacteria	Patients, No.	Male/Female, No.	% Positive
Escherichia coli	115	37/78	62.5
Klebsiella pneumonia	38	16/22	20.7
Pseudomonas aeruginosa	17	7/10	9.2
Streptococcus agalactiae	6	2/4	3.3
Enterococcus	3	1/2	1.6
Acinetobacter baumannii	2	1/1	1.1
Candida albicans	2	0/2	1.1
Staphylococcus xylosus	1	1/0	0.5

positive, FP represents false positive, and FN represents false negative. Its value ranges from 0 to 1, with higher values indicating greater similarity between the prediction and the true label. The Jaccard index compares the similarity between 2 sets by calculating the proportion of shared elements using the formula TP / (TP + FP + FN). Originally designed to evaluate the accuracy of image segmentation algorithms, the core concept of measuring similarity or overlap between 2 sets suggests that these 2 metrics may have valuable applications in a wider range of contexts.

Subsequently, we employed the DeLong test to determine significant differences in AUC among various models, conducting a comprehensive evaluation using other relevant indicators [23]. The DeLong test is a statistical methodology employed to assess the significance of the difference between 2 correlated areas under the AUC. Notably, this test does not directly rely on the standard deviation of the data but instead leverages crucial statistics derived from ROC analysis, which encapsulate the ordering and quantity of predictions across distinct classes (positive and negative). Specifically, the DeLong test considers the variances (var_A, var_B) of the AUC corresponding to each ROC curve, as well as the covariance (covar_AB) between them. Utilizing this information, it computes a *z*-score, thereby evaluating the statistical significance of the difference between the 2 AUCs. This process facilitated the identification of the optimal classifier model that effectively integrates both radiomics and clinical features. Furthermore, we developed a multimodal combined model based on these characteristics. By reevaluating and validating the model's performance, we obtained corresponding results for all subject characteristic curves. Finally, we quantified the net benefit of different threshold probabilities in the training and testing cohorts using decision curve analysis (DCA) to assess the clinical utility of the model in kidney stone classification [24]. Additionally, we employed a calibration curve to evaluate the model's calibration status.

Statistical Analysis

Statistical analysis was performed using Python (version 3.10.1; http://www.python.org) and SPSS (version 26.0, IBM) software.

Continuous variables were compared using *t* tests and expressed as mean ± standard deviation based on their distribution. Categorical variables were compared using the χ^2 test or Fisher exact test and presented as numbers (percentages). Univariate and multivariate LR analyses were conducted to detect clinical indicators and identify risk factors. Variables with *P* values <.05 in univariate analysis were entered into multivariate LR analysis. Two-tailed *P* values <.05 were considered statistically significant.

Comparative Analysis

In accordance with established comparative analysis methodologies, we conducted a comprehensive literature search using predefined keywords, including "radiomics" and "infection related to kidney stones." Despite the exhaustive nature of our search, we identified only a limited number of studies meeting the inclusion criteria, with 1 study closely aligning with our research focus [11]. To ensure a robust and systematic comparison, we meticulously designed a detailed comparison table that highlights the key similarities and distinctions between our findings and those of the included study. The comparative analysis primarily encompasses 3 distinct aspects. First, we conducted a meticulous comparison of the foundational information across various studies, encompassing but not limited to the first author's identity, country, classification criteria, study

design, sample size for diagnostic accuracy, mean or median age of participants, and the year of publication. Second, we employed the Quality Assessment of Studies for Diagnostic Accuracy (QUADAS-2) guidelines to scrutinize each study [25]. By systematically addressing the signaling questions outlined in the QUADAS-2 tool with responses of "yes," "no," or "unclear," we conducted a nuanced evaluation of the potential bias risks and methodological quality of each study. Furthermore, in the third aspect of our analysis, we zeroed in on the performance metrics of the models under investigation. We focused on pivotal indicators such as the AUC, sensitivity, specificity, accuracy, Acc_mean, and Acc_std, and paid particular attention to the application of statistical methods like LR.

RESULTS

Patient Population and Radiological Characteristics

This retrospective study included 461 patients (218 men and 243 women), including 322 patients in the training set and 139 patients in the test set. Table 1 presents the statistical data of positive urine cultures in patients with kidney stones. Table 2 provides a detailed summary of the patients' clinical characteristics, revealing no significant differences between the training and test sets. Notably, there were significant differences in clinical characteristics (eg, urine white blood cell

Table 2.   Clinical Characteristics in the Training and Test Cohorts

Variable	Training Cohort (n = 322)			Testing Cohort (n = 139)		
	Uninfected (n = 192)	Infected (n = 130)	*P* Value	Uninfected (n = 83)	Infected (n = 56)	*P* Value
Age	50.63 ± 13.271	51.37 ± 12.938	.621	52.21 ± 14.249	49.77 ± 12.267	.315
White blood cell count	133.13 ± 689.061	539.71 ± 1151.792	<.001*	155.99 ± 878.974	593.11 ± 1130.171	.017*
Urine pH	6.013 ± 0.4355	6.565 ± 0.6774	<.001*	6.00 ± 0.4814	6.598 ± 0.5674	<.001*
Urea nitrogen	5.5366 ± 1.8596	6.7908 ± 3.4805	<.001*	5.8747 ± 2.8169	6.9789 ± 3.4869	.042*
Uric acid	345.78 ± 93.722	367.27 ± 116.034	.080	347.31 ± 98.061	396.25 ± 137.765	.024*
Serum creatinine	86.159 ± 36.6128	104.582 ± 62.5254	.003*	86.293 ± 36.1133	98.007 ± 50.2553	.112
Sex, No. (%)						
Female	85 (44.27)	67 (51.34)	.212	41 (49.40)	25 (44.64)	.607
Male	107 (55.73)	63 (48.66)		42 (50.60)	31 (55.36)	
Urinary nitrite, No. (%)						
Negative	122 (63.54)	47 (36.15)	<.001*	52 (62.65)	21 (37.5)	.006*
Positive	64 (33.33)	69 (53.07)		30 (36.14)	31 (55.36)	
Strongly positive	6 (3.13)	14 (10.78)		1 (1.21)	4 (7.14)	
Leukocyte esterase, No. (%)						
Negative	137 (71.35)	52 (40)	<.001*	52 (62.65)	17 (30.36)	<.001*
Positive	47 (24.48)	65 (50)		30 (36.14)	28 (50)	
Strongly positive	8 (4.17)	13 (10)		1 (1.21)	11 (19.64)	
Infrared spectrum, No. (%)						
CM	139 (72.40)	20 (15.38)	<.001*	52 (62.65)	10 (17.86)	<.001*
CD	33 (17.19)	10 (7.69)		24 (28.92)	9 (16.07)	
CA	11 (5.73)	61 (46.93)		5 (6.02)	29 (51.89)	
AU	9 (4.68)	39 (30)		2 (2.41)	8 (14.28)	

Data are presented as mean ± standard deviation unless otherwise indicated.

Abbreviations: AU, anhydrous uric acid; CA, carbonated apatite; CD, calcium oxalate dihydrate; CM, calcium oxalate monohydrate.

*P < .05.

Table 3. Univariate and Multivariable Logistic Regression Analyses for Selecting Clinical Features of Model Development

Variable	Univariate Analysis				Multivariate Analysis			
	Coefficient	SE	OR (95% CI)	P Value	Coefficient	SE	OR (95% CI)	P Value
Age	−0.0075	0.002	0.99 (−.011, −.004)	<.001*	−0.0327	0.009	3.27 (−.051, −.015)	<.001*
Sex	−0.4607	0.132	0.63 (−.719, −.202)	<.001*	−0.4381	0.243	1.0 (−.915, .039)	.072
Infrared spectrum	0.1083	0.041	1.11 (.027–.189)	.009*	1.1862	0.125	2.54 (.942–1.431)	<.001*
White blood cell count	0.0004	0.000	1.0 (.000–.001)	.007*	0.0004	0.000	1.0 (6.74e-05, .001)	.019*
Urinary nitrite	0.1916	0.118	1.21 (−.039, .422)	.104	…	…	…	
Leukocyte esterase	0.3102	0.118	1.36 (.078–.542)	.009*	0.9311	0.194	0.75 (.551–1.311)	<.001*
Serum creatinine	−0.0019	0.001	1.0 (−.004, −8.56e-05)	.040*	0.0039	0.003	0.65 (−.002, .083)	.159
Urine pH	−0.0478	0.015	0.95 (−.077, −.018)	.001*	−0.2822	0.102	1.0 (−.482, −.093)	.006*
Urea nitrogen	−0.0266	0.014	0.97 (−.054, .001)	.057	…	…	…	
Uric acid	−0.0008	0.000	1.0 (−.001, −.000)	.002*	−0.0003	0.001	0.97 (−.002, .002)	.795

Abbreviations: CI, confidence interval; OR, odds ratio; SE, standard error.

*$P < .05$.

count, urine nitrite content, leukocyte esterase content, urine pH, and urea nitrogen content) between patients with and without infection in both the training and test sets. Univariate and multivariate LR analyses demonstrated that these clinical characteristics were independent predictive factors significantly associated with the presence of infection in patients with kidney stones (Table 3). Therefore, we incorporated these indicators as clinical features for subsequent classification research.

Radiomics Features Models and Performances

We extracted 414 radiomics features from the kidney stone ROI in each patient. Using the Mann-Whitney U test and LASSO regression feature selection methods, we reduced the feature set and selected 18 features with nonzero coefficients. These features comprise 6 shape features, 4 first-order statistics (FOS) features, 2 gray level co-occurrence matrix (GLCM) features, 2 gray level size zone matrix (GLSZM) features, 3 gray level dependence matrix (GLDM) features, and 1 gray level run Length Matrix (GLRLM) feature (Figure 3, Table 4). In subsequent classifier model studies focusing on radiomics features, the MLP model demonstrated exceptional and consistent classification performance on both training and testing datasets in identifying KS-UTIs patients. Specifically, on the training dataset, the AUC, accuracy, Acc_mean, Acc_std, sensitivity, specificity, PPV, NPV, DSC, and Jaccard index were 0.96, 0.91, 0.84, 0.06, 0.88, 0.92, 0.88, 0.92, 0.88, and 0.79, respectively. On the testing dataset, the corresponding values were 0.94, 0.87, 0.84, 0.02, 0.86, 0.88, 0.83, 0.90, 0.84, and 0.73, respectively (Figure 4A and 4B, Table 5). The results of the DeLong test also indicated significant differences between the MLP model and other models (Supplementary Tables 1 and 2). Although LR and other deep learning models exhibited strong classification performance, they were still inferior to MLP in performance evaluation. Certain shallow models, such as RF and XGBoost, showed excellent classification performance on the training dataset but performed less satisfactorily on the testing dataset, indicating overfitting.

Clinical Features Models and Performances

In the screening process for clinical feature classification models, we also compared model performance metrics on both the training and testing datasets and conducted the DeLong test (Supplementary Tables 3 and 4). The MLP model demonstrated the most superior classification performance across both datasets. Specifically, on the training dataset, the AUC, accuracy, Acc_mean, Acc_std, sensitivity, specificity, PPV, NPV, DSC, and Jaccard index were 0.95, 0.89, 0.84, 0.07, 0.84, 0.93, 0.89, 0.89, 0.86, and 0.76, respectively. On the testing dataset, the corresponding values were 0.91, 0.88, 0.85, 0.02, 0.80, 0.93, 0.88, 0.88, 0.84, and 0.73, respectively (Figure 4C and 4D, Table 5). Although LR and other deep learning models performed slightly less well than MLP in the performance evaluation, they still exhibited excellent results in identifying KS-UTIs patients. Among the shallow learning models, RF and XGBoost also showed signs of overfitting, as their exceptional performance on the training dataset was not reflected in the testing dataset. Based on a comprehensive and objective evaluation, combined with previous research on radiomics features, we ultimately selected the MLP model as the optimal classification model for subsequent studies.

Combined Features Model Construction and Validation

By combining the predictive factors from the previously obtained clinical features with the 18 radiomic features, we developed a combined model based on the MLP approach. We conducted a unified and objective evaluation of the MLP model results for radiomic features, clinical features, and the combined model. The ROC curves, DCA curves, and calibration curves for the subjects in the training and testing datasets for these 3 models are shown in Figure 5. The results of the DeLong test are presented in Supplementary Tables 5 and 6. In both the training and testing datasets, the combined model
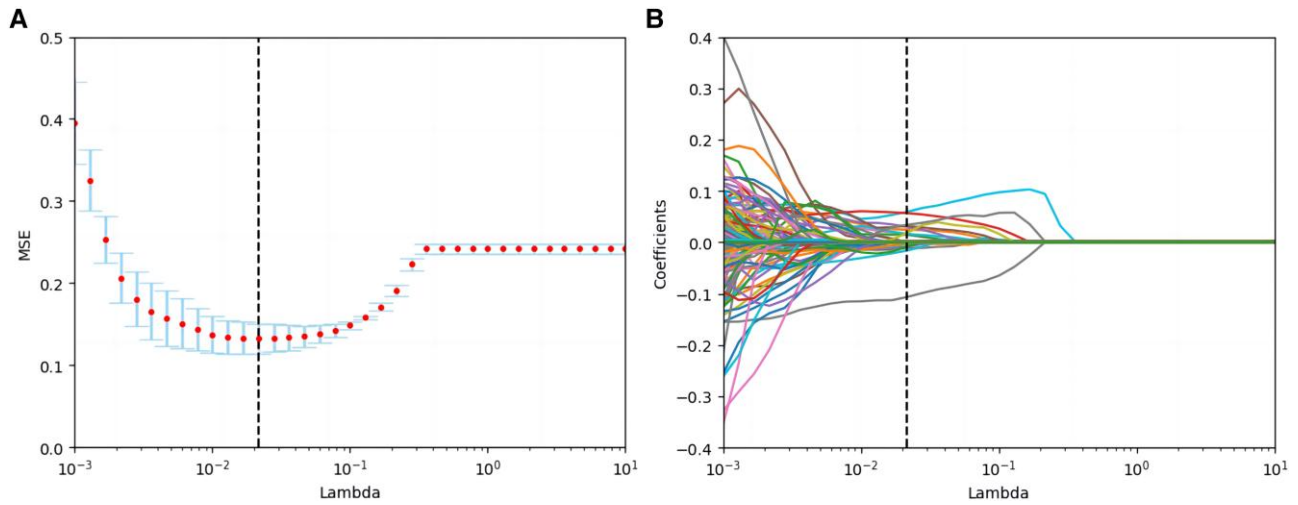
Figure 3. Feature selection with the least absolute shrinkage and selection operator (LASSO) regression model. *A,* The LASSO model's tuning parameter (λ) selection used 5-fold cross-validation via minimum criterion. The vertical lines indicate the optimal value of the LASSO tuning parameter (λ). *B,* LASSO coefficient profile plot with different log (λ) was shown. The vertical dashed lines represent all radiomics features with nonzero coefficients selected with the optimal λ value. Abbreviation:MSE, mean square error.

demonstrated superior discriminatory ability compared to the clinical and radiomic feature models. Specifically, in the training dataset, the AUC, accuracy, Acc_mean, Acc_std, sensitivity, specificity, PPV, NPV, DSC, and Jaccard index for the combined model were 0.98, 0.94, 0.90, 0.02, 0.91, 0.96, 0.94, 0.94, 0.93, and 0.86, respectively. In the testing dataset, the corresponding values were 0.95, 0.92, 0.92, 0.01, 0.89, 0.94, 0.91, 0.93, 0.90, and 0.82 (Table 5). Additionally, the DCA results indicated that the combined model provided the most significant net benefit in classifying infectious stones in both the training and testing datasets. The calibration curve results showed that the combined model exhibited good agreement.

Comparative Analysis With Similar Articles

The comparative analysis with similar articles is shown in Table 6; our study stands out from previous similar endeavors by adopting more advanced learning methodologies, particularly deep learning techniques. Notably, we observed a substantial improvement in the recognition performance for KS-UTIs when utilizing these advanced methods.

DISCUSSION

In this study, we retrospectively compared the classification performance of various machine learning classifiers, including 3 shallow learning algorithms and 4 deep learning algorithms, with the aim of developing a classifier capable of effectively identifying KS-UTIs. We found that radiomics features and clinical features based on the MLP model performed well, demonstrating a good fit between the training and testing datasets. Specifically, the radiomics features yielded an AUC of 0.96 for the training dataset and 0.94 for the testing dataset, while the

Table 4. Radiomics Features' Selection

Feature Class	Feature Name	Feature Quantity
shape	LeastAxisLength	1
	Maximum2DDiameterColumn	1
	Maximum2DDiameterSlice	1
	MinorAxisLength	1
	Sphericity	1
	SurfaceVolumeRatio	1
firstorder	Minimum	2
	Range	1
	Skewness	1
GLCM	Idn	1
	MCC	1
GLSZM	LargeAreaHighGrayLevelEmphasis	1
	ZoneEntropy	1
GLDM	DependenceNonUniformityNormalized	1
	LargeDependenceHighGrayLevelEmphasis	1
	DependenceEntropy	1
GLRLM	RunVariance	1

Abbreviations: GLCM, gray level co-occurrence matrix; GLDM, gray level dependency matrix; GLRLM, gray level run length matrix; GLSZM, gray level size zone matrix; MCC, maximum correlation coefficient.

clinical indicator features achieved an AUC of 0.95 for the training dataset and 0.91 for the testing dataset. Furthermore, by combining these radiomics and clinical features, we developed a noninvasive combined model based on CT imaging. This model significantly improved prediction performance compared to the single radiomics and clinical feature model in distinguishing whether stones are associated with infection, with an AUC of 0.98 for the training dataset and 0.95 for the testing dataset. This model provides a feasible and reliable noninvasive method for the future diagnosis of KS-UTIs.
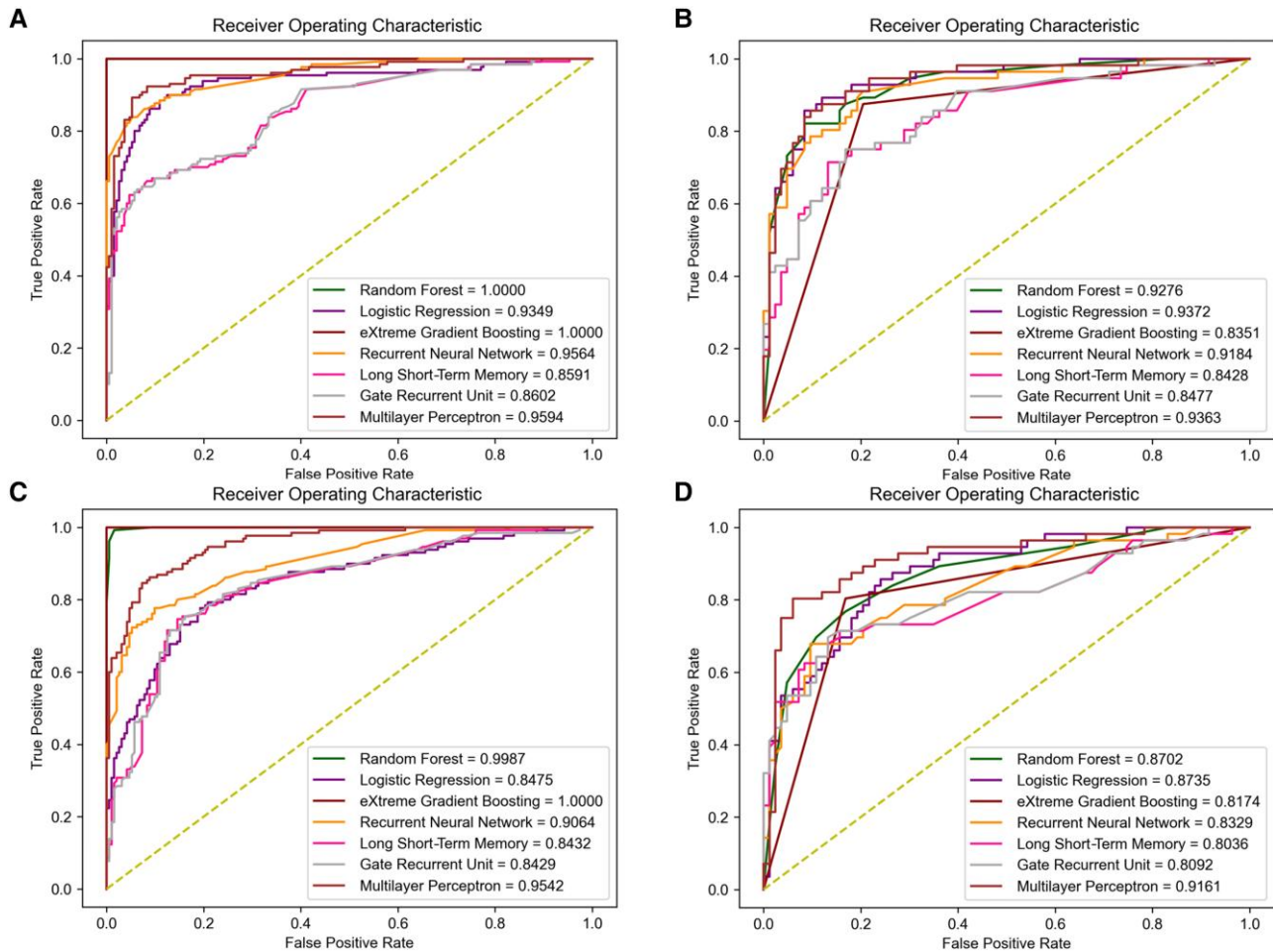
Figure 4. The fitting performance of different features across all models. *A*, Receiver operating characteristic (ROC) curves of all models when fitting the radiomics feature training set. *B*, ROC curves of all models when fitting the radiomics feature test set. *C*, ROC curves of all models when fitting the clinical feature training set. *D*, ROC curves of all models when fitting the clinical feature test set.

LASSO regression as the feature selection method in this study has demonstrated its reliability through an abundance of prior research, lending credibility to its application within the current investigation. In LASSO regression, feature importance is indirectly reflected by the coefficients assigned to each feature. Larger absolute coefficient values indicate greater importance in model construction. However, due to the regularization mechanism, these coefficients represent the contribution of features to predictions under a specific regularization strength, rather than their inherent importance in the original dataset. Through LASSO regression's dimensionality reduction, we identified key features with nonzero or large coefficients that significantly improve predictive accuracy. These features enhance the model in several ways: They increase prediction accuracy by maintaining strong associations with the target variable, simplify the model by removing redundant information, and improve interpretability [26]. Additionally, LASSO regression reduces computational complexity, making it especially useful for large-scale or real-time datasets [27].

Last, by focusing on essential features, LASSO regression enhances model stability, maintaining robust predictions even with minor data variations, thereby providing a more reliable basis for decision-making.

We previously found that few studies have compared the performances of various classifiers in terms of predicting whether kidney stones are accompanied by urinary tract infection [11]. Thus, we compared performance among machine learning classifiers to address this lack of information. The results of this study indicate that the MLP model outperforms other machine learning models in terms of various classification performance metrics. This proven algorithm demonstrates excellent classification accuracy among existing algorithms. The MLP model processes input data through multiple fully connected layers, allowing it to learn and simulate complex nonlinear relationships. Due to its ability to avoid overfitting, the MLP model is widely used in classification tasks across various domains [28]. On the other hand, RF and XGBoost demonstrated similar performance. They both performed well in the training set but

Table 5. Diagnostic Performance of Different Models in Predicting Concomitant Infection With Kidney Stones in the Training and Testing Cohorts

Feature Type	Model	Set	AUC	Acc	Acc_mean	Acc_std	Sensitivity	Specificity	PPV	NPV	DSC	Jaccard Index
Radiomics features	RF	Training	1.00	1.00	0.87	0.13	1.00	1.00	1.00	1.00	1.00	1.00
		Test	0.93	0.87	0.85	0.03	0.88	0.87	0.82	0.91	0.84	0.73
	LR	Training	0.93	0.88	0.85	0.04	0.86	0.90	0.85	0.91	0.85	0.75
		Test	0.94	0.88	0.84	0.06	0.89	0.87	0.82	0.92	0.85	0.75
	XGBoost	Training	1.00	1.00	0.80	0.09	1.00	1.00	1.00	1.00	1.00	1.00
		Test	0.84	0.83	0.88	0.04	0.88	0.80	0.74	0.90	0.80	0.67
	RNN	Training	0.96	0.90	0.82	0.03	0.85	0.94	0.90	0.90	0.87	0.77
		Test	0.92	0.85	0.82	0.02	0.80	0.88	0.82	0.87	0.81	0.68
	LSTM	Training	0.86	0.81	0.75	0.02	0.62	0.94	0.87	0.79	0.73	0.57
		Test	0.84	0.79	0.79	0.02	0.62	0.90	0.81	0.78	0.71	0.55
	GRU	Training	0.86	0.77	0.74	0.04	0.53	0.93	0.84	0.75	0.65	0.48
		Test	0.85	0.74	0.77	0.01	0.45	0.94	0.83	0.72	0.58	0.41
	MLP[a]	Training	0.96	0.91	0.84	0.06	0.88	0.92	0.88	0.92	0.88	0.79
		Test	0.94	0.87	0.84	0.02	0.86	0.88	0.83	0.90	0.84	0.73
Clinical features	RF	Training	0.99	0.98	0.79	0.03	0.97	0.99	0.99	0.98	0.98	0.96
		Test	0.87	0.77	0.76	0.08	0.75	0.78	0.70	0.82	0.72	0.57
	LR	Training	0.85	0.78	0.77	0.05	0.78	0.78	0.71	0.84	0.74	0.59
		Test	0.87	0.78	0.75	0.05	0.86	0.73	0.69	0.88	0.76	0.62
	XGBoost	Training	1.00	1.00	0.73	0.21	1.00	1.00	1.00	1.00	1.00	1.00
		Test	0.82	0.82	0.69	0.13	0.80	0.83	0.76	0.86	0.78	0.64
	RNN	Training	0.91	0.86	0.79	0.06	0.76	0.93	0.88	0.85	0.81	0.69
		Test	0.83	0.78	0.81	0.03	0.62	0.89	0.80	0.78	0.70	0.54
	LSTM	Training	0.84	0.79	0.74	0.07	0.76	0.80	0.72	0.83	0.74	0.59
		Test	0.80	0.78	0.79	0.01	0.70	0.84	0.75	0.80	0.72	0.57
	GRU	Training	0.84	0.81	0.75	0.11	0.72	0.87	0.79	0.82	0.75	0.60
		Test	0.80	0.78	0.75	0.04	0.62	0.89	0.80	0.78	0.70	0.54
	MLP[a]	Training	0.95	0.89	0.84	0.07	0.84	0.93	0.89	0.89	0.86	0.76
		Test	0.91	0.88	0.85	0.02	0.80	0.93	0.88	0.88	0.84	0.73
Combined features	MLP[a]	Training	0.98	0.94	0.90	0.02	0.91	0.96	0.94	0.94	0.93	0.86
		Test	0.95	0.92	0.92	0.01	0.89	0.94	0.91	0.93	0.90	0.82

Abbreviations: Acc, accuracy; AUC, area under the curve; DSC, dice similarity coefficient; GRU, gate recurrent unit; LR, logistic regression; LSTM, long short-term memory; MLP, multilayer perceptron; NPV, negative predictive value; PPV, positive predictive value; RF, random forest; RNN, recurrent neural network; std, standard deviation; XGBoost, extreme gradient boosting.
[a]Classifier model with the best classification performance.

displayed poor performance in the test set. We hypothesize that this overfitting phenomenon may be related to the underlying logic of these 2 scripts. Both RF and XGBoost are ensemble models based on decision trees, capable of constructing highly complex decision boundaries. However, in certain instances, this high complexity may also lead to overfitting to the nuances of the training data [29, 30]. Moreover, we found that although LR, RNN, LSTM, and GRU had lower AUCs in both the training and test sets (compared with MLP), it continued to exhibit excellent classification accuracy, indicating that they could be suitable for classification of KS-UTI patients. In summary, these results demonstrate the significant potential of machine learning in medical fields. They validate our initial hypothesis that machine learning models can effectively utilize radiomics and clinical features to assist in distinguishing whether kidney stones are accompanied by infection.

We have discovered significant differences in radiomics features, primarily composed of shape features and FOS features, when distinguishing whether kidney stone patients have KS-UTIs. Shape features describe the morphological attributes of the ROI, such as volume, surface area, and compactness. These features are intuitive and directly related to the growth patterns and aggressiveness of kidney stones [31]. On the other hand, FOS features are primarily based on histogram analysis of the image, reflecting the distribution of pixel intensities, including mean, median, and standard deviation [32]. These features are relatively simple to calculate and can capture fundamental intensity information from the image, indicating that radiomics features can detect potential stone heterogeneity that is not visible to the naked eye. Although second-order and higher-order texture features account for a smaller proportion, they still play a crucial role in image processing and analysis. By quantifying the distribution patterns of pixel or voxel intensities in an image, these features provide insightful information about the local structure and patterns [33–35]. These findings contribute to the application of radiomics in the study of kidney stones with infection and offer new perspectives for addressing other clinical issues.
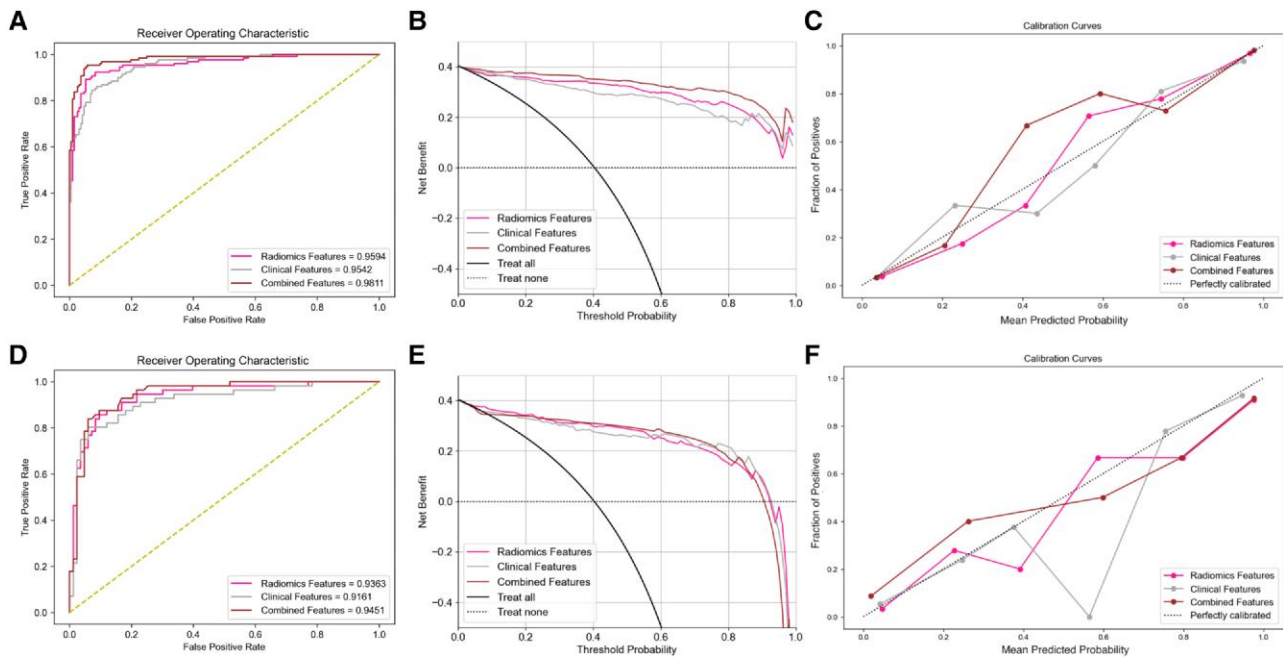
Figure 5. The performance between the combined model and single-feature models. *A*, The receiver operating characteristic (ROC) curves for the training set of 3 models based on multilayer perceptron (MLP): the radiomics feature model, the clinical feature model, and the combined feature model. *B*, The decision curve analysis (DCA) curves were generated for the same 3 models using the training set. *C*, The calibration curves were created for the same 3 models using the training set. *D*, The ROC curves for the test set of 3 models based on MLP: the radiomics feature model, the clinical feature model, and the combined feature model. *E*, The DCA curves were generated for the same 3 models using the test set. *F*, The calibration curves were created for the same 3 models using the test set. Abbreviations: DCA, decision curve analysis; MLP, multilayer perceptron; ROC, receiver operating characteristic.

Although radiomics features demonstrated superior predictive power compared to clinical features in this study, the importance of clinical feature information cannot be overlooked and still plays a crucial role. Our univariate and multivariate LR analyses revealed that clinical indicators can also serve as effective predictors for determining whether kidney stones are KS-UTIs. In a study of urine culture samples, we found that some bacteria play a dominant role in infection, including *Escherichia coli* and *Klebsiella pneumoniae*; these members of Enterobacteriaceae have been identified in previous work [36, 37]. Their presence during the course of disease has a considerable impact on prognosis of KS-UTIs patients; differences in treatment regimen can also affect patients' outcomes [38, 39]. Simultaneously, each type of kidney stone has a distinct composition, and we speculated that radiomics features could reflect this heterogeneity. The present study confirmed our hypothesis, revealing that uric acid and carbonate apatite were the main components of infection-associated kidney stones; monohydrate calcium oxalate and dihydrate calcium oxalate were the main components of non-infection-associated kidney stones. These findings suggest that stone composition is also an effective predictor of whether kidney stones are accompanied by infection. Uric acid and carbonate apatite are complex crystal aggregates that accumulate in organic matrices, contributing to 10%–15% of urinary system stone diseases [40]. Additionally,

increase of urine pH and urease content can effectively predict the incidence of kidney stone infection; while urine is acidic (pH 6.8), carbonate apatite begins to crystallize, whereas struvite begins to precipitate while urine is alkaline (pH 7.2) [11], consistent with our findings. Furthermore, patients with uric acid stones are older, have higher body mass index and lower urine pH, and are more likely to be women, compared with patients with calcium stones [41]. These findings suggest that the occurrence of kidney stones with infection may not solely result from a chain reaction triggered by specific proteins or genes. Multiple factors, including environmental conditions, lifestyle habits, dietary practices, and the patient's immune system status, may also play significant roles in the development of the disease [42]. This simultaneously validates our second hypothesis, namely, that radiomics and clinical features can assist in elucidating the formation mechanisms of the condition.

In addition, we discovered that the combined model developed using radiomics features and clinical features demonstrated optimal classification performance in our results, with an AUC of 0.988 in the training set and 0.95 in the test set. This finding not only highlights the excellent classification capability of the combined model but also underscores the importance of multimodal data integration in disease diagnosis. By comprehensively utilizing both radiomics and clinical features, we can more effectively capture the complexity and diversity of

Table 6. Comparative Analysis With Similar Studies

References	Country	Classification Criteria	Study Design	Sample Size for Diagnostic Accuracy	No. of Patients (% Female)	Imaging Modality	Mean or Median Age	AI Model (per-Patient/per-Node Diagnostic Output)	Year	Reference Standard
Zheng et all[11]	China	KS-UTIs	Retrospective single-center	884	426 (35.5%)	CT	42.68	Radiomics (per-patient)	2021	Radiology
This research	China	KS-UTIs	Retrospective multicenter	139	218 (47.3%)	CT	56.53	Radiomics + deep learning (per-patient)	2024	Radiology

Risk of Bias

	Patient Selection			Index Test			Reference Standard	Feature Selection
	Was the statistical management adequate?	Were the inclusion/exclusion criteria specified?	Was the type of study (retrospective or prospective) specified?	Were the imaging acquisition protocol and the segmentation method(s) detailed?	Was the image processing approach detailed?	Was the validation independent (ie, not internal)?	Was the reference standard adequate?	Number of feature selection methods
Zheng et all[11]	Yes	Yes	Yes	Yes	Yes	No	Yes	4
This research	Yes	Yes	Yes	Yes	Yes	No	Yes	1

Applicability Concerns

	Patient Selection	Index Test	Reference Standard
Zheng et all[11]	Yes	Yes	Yes
This research	Yes	Yes	Yes

	Sensitivity	Specificity	Accuracy	Accuracy_Mean	Accuracy_Standard Deviation	AUC	Extracted Radiomics Features		DCA	Calibration Curves	DeLong Test
Zheng et all[11]	Unclear	Nnclear	Nnclear	Nnclear	Nnclear	0.86	A, B, C, D, E	Logistic Regression	Yes	Yes	Yes
This research	0.91	0.96	0.94	0.90	0.02	0.98	A, B, C, D	Logistic Regression	Yes	Yes	Yes

Abbreviation: A, first-order statistics features; AI, artificial intelligence; AUC, area under the curve; B, shape-based features; C, statistics-based textural features; CT, computed tomography; D, Laplacian of Gaussian–filtered features; DCA, decision curve analysis; E, wavelet features; KS-UTI, kidney stone–associated urinary tract infection.

diseases, thereby enhancing diagnostic accuracy and reliability [43]. Furthermore, through a meticulous examination of the relevant performance metrics across various classification models, we were pleased to observe that the optimal model, identified by key indicators such as accuracy, Acc_mean, and Acc_std, aligned perfectly with the conclusions drawn from the DeLong test. The DeLong test, a widely recognized nonparametric statistical method prevalent in medical statistics and other fields, is renowned for its rigor and precision. This remarkable concordance not only affirms the scientific rigor and rationale behind our model selection process but also highlights the strong potential and reliability of these evaluation metrics in accurately assessing model performance. DCA results indicate that the combined model significantly improves the effectiveness of clinical decision-making compared to using radiomics or clinical features alone. This suggests that the model can precisely guide treatment, optimize resource allocation, and enhance KS-UTIs patients' diagnosis and treatment experiences, ultimately maximizing clinical net benefits. Furthermore, the excellent performance of the combined model on the calibration curve demonstrates its outstanding calibration accuracy. This signifies that the model accurately converts predicted probabilities into the likelihood of actual event occurrences, providing decision-makers with reliable probability estimates [44]. Last, these results validate our third hypothesis: that combining radiomics features with clinical features can improve classification accuracy.

Finally, the outcomes of this comparative analysis hold profound value and significance. By meticulously comparing with similar studies across 3 dimensions—basic information, research quality assessment, and model performance evaluation—we have not only delineated the temporal-spatial contexts and foundational frameworks of each study but also illuminated their unique strengths and potential limitations in design and execution. This process has not only deepened our understanding of the current research landscape in KS-UTI identification but also laid a solid foundation for our subsequent research endeavors. Notably, the results of this comparative analysis unequivocally demonstrate the superior performance of deep learning models over radiomics models in the task of KS-UTIs identification. This finding not only validates the robust capabilities of deep learning technologies in processing complex medical imaging data but also attests to the scientific rigor and rationality of our study in terms of methodology selection, model construction, and optimization. It further reinforces our confidence in the promising application prospects of deep learning in medical diagnosis and inspires us to continually explore more innovative solutions to address increasingly intricate medical challenges.

This study has several limitations. First, it is a retrospective, single-center study. Despite the use of strict inclusion and exclusion criteria, there may be potential selection bias, which

we aim to address through external validation in future studies. Second, defining the boundaries of manual segmentation involves some controversy and subjectivity. We hope to automate this process in the future using deep learning–based automatic segmentation. Additionally, our study employed 2 classic feature selection methods: the Mann-Whitney U test and LASSO regression. Both methods have inherent limitations. The Mann-Whitney U test, based on nonparametric ranking properties, may not be suitable for nonnormal distributions or small sample sizes. It is also not sensitive to ordinal categorical variables and cannot handle multivariate relationships. LASSO regression, on the other hand, has a strong dependence on regularization parameters and can be unstable when dealing with highly correlated features. Moreover, it may sacrifice some predictive accuracy in pursuit of model sparsity. In future studies, we intend to explore and validate alternative feature selection methods to overcome these limitations.

CONCLUSIONS

In summary, our study has demonstrated that both radiomics feature and clinical feature models can accurately distinguish whether kidney stones are KS-UTIs across multiple classifiers. Additionally, we developed a multimodal combined model that effectively enhances the classification accuracy for patients with KS-UTIs. The findings of this study provide new insights and methodologies for future disease diagnosis, which are expected to drive further advancements in medical diagnostic technology.

References

1. Ming S, Tian J, Ma K, et al. Oxalate-induced apoptosis through ERS-ROS-NF-κB signalling pathway in renal tubular epithelial cell. Mol Med **2022**; 28:88.
2. Gambaro G, Vezzoli G, Casari G, et al. Genetics of hypercalciuria and calcium nephrolithiasis: from the rare monogenic to the common polygenic forms. Am J Kidney Dis **2004**; 44:963–86.
3. Cohen TD, Preminger GM. Struvite calculi. Semin Nephrol **1996**; 16:425–34.
4. Shen C, Zhu Q, Dong F, et al. Identifying two novel clusters in calcium oxalate stones with urinary tract infection using 16S rDNA sequencing. Front Cell Infect Microbiol **2021**; 11:723781.
5. Peerapen P, Thongboonkerd V. Kidney stone prevention. Adv Nutr **2023**; 14: 555–69.
6. Moe OW. Kidney stones: pathophysiology and medical management. Lancet **2006**; 367:333–44.
7. Antonelli JA, Maalouf NM, Pearle MS, Lotan Y. Use of the National Health and Nutrition Examination Survey to calculate the impact of obesity and diabetes on cost and prevalence of urolithiasis in 2030. Eur Urol **2014**; 66:724–9.
8. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology **2016**; 278:563–77.
9. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol **2017**; 14:749–62.
10. Liu S, Chen X, Lin T. Lymphatic metastasis of bladder cancer: molecular mechanisms, diagnosis and targeted therapy. Cancer Lett **2021**; 505:13–23.
11. Zheng J, Yu H, Batur J, 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**; 100:870–80.
12. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med **2019**; 25:24–9.
13. Wang X, Fan J. Spatiotemporal molecular imaging is a critical part of spatiotemporal molecular medicine. Clin Transl Med **2021**; 11:e347.
14. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology **2020**; 295:328–38.
15. Rinaldi L, De Angelis SP, Raimondi S, et al. Reproducibility of radiomic features in CT images of NSCLC patients: an integrative analysis on the impact of acquisition and reconstruction parameters. Eur Radiol Exp **2022**; 6:2.
16. Tagare HD, DeLorenzo C, Chelikani S, Saperstein L, Fulbright RK. Voxel-based logistic analysis of PPMI control and Parkinson's disease DaTscans. Neuroimage **2017**; 152:299–311.
17. Menze BH, Kelm MB, Masuch R, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics **2009**; 10:213.
18. Sushentsev N, Rundo L, Abrego L, et al. Time series radiomics for the prediction of prostate cancer progression in patients on active surveillance. Eur Radiol **2023**; 33:3792–800.
19. He QH, Feng JJ, Wu LC, et al. Deep learning system for malignancy risk prediction in cystic renal lesions: a multicenter study. Insights Imaging **2024**; 15:121.
20. Yun J, Park JE, Lee H, et al. Radiomic features and multilayer perceptron network classifier: a robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. Sci Rep **2019**; 9:5746.
21. Zeiler MD. ADADELTA: an adaptive learning rate method. arXiv [Preprint]. Posted online 22 December **2012**. doi:10.48550/arXiv.1212.5701
22. Nai YH, Teo BW, Tan NL, et al. Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset. Comput Biol Med **2021**; 134: 104497.
23. 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.
24. Yu X, Zhou L, Yang C. A prospective validation cohort study of a prediction model on non-sentinel lymph node involvement in early breast cancer. Ann Surg Oncol **2019**; 37(15 Suppl):559.
25. Bouwmeester W, Zuithoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med **2012**; 9:1–12.
26. Gui J, Li H. Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. Bioinformatics **2005**; 21:3001–8.
27. Hu S, Lyu X, Li W, et al. Radiomics analysis on noncontrast CT for distinguishing hepatic hemangioma (HH) and hepatocellular carcinoma (HCC). Contrast Media Mol Imaging **2022**; 2022:7693631.
28. Feng J, Zeng R, Geng Y, et al. Automatic differentiation of ruptured and unruptured intracranial aneurysms on computed tomography angiography based on deep learning and radiomics. Insights Imaging **2023**; 14:76.
29. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Knowledge Discovery and Data Mining, San Francisco, CA, 13–17 **August 2016**.
30. Oshiro TM, Perez PS, Baranauskas JA. How many trees in a random forest? In: The 8th international conference on Machine Learning and Data Mining in Pattern Recognition. Berlin, Germany, **2012**.
31. Yap FY, Varghese BA, Cen SY, et al. Shape and texture-based radiomics signature on CT effectively discriminates benign from malignant renal masses. Eur Radiol **2021**; 31:1011–21.
32. Santo BA, Jenkins TJD, Ciecierska SSK, et al. MicroCT and histological analysis of clot composition in acute ischemic stroke—a comparative study of MT-retrieved clots and clot analogs. Clin Neuroradiol **2024**; 34:431–9.
33. Kairuddin WNHW, Mahmud WMHW. Texture feature analysis for different resolution level of kidney ultrasound images. In: International Research and Innovation Summit. Melaka, Malaysia, **2017**; 012136.
34. Khalil MM. Radiomics feature selection in 18F-FDG PET imaging: investigation of the most robust and reproducible candidates. Egypt J Nucl Med **2020**; 21: 80–102.
35. Thibault G, Fertil B, Navarro C, Pereira S, Mari JL. 10th International Conference on Pattern Recognition and Information Processing. Minsk, Belarus, **2009**; p.140–145.
36. Tavichakorntrakool R, Prasongwattana V, Sungkeeree S, et al. Extensive characterizations of bacteria isolated from catheterized urine and stone matrices in patients with nephrolithiasis. Nephrol Dial Transplant **2012**; 27:4125–30.
37. Chen D, Jiang C, Liang X, et al. Early and rapid prediction of postoperative infections following percutaneous nephrolithotomy in patients with complex kidney stones. BJU Int **2019**; 123:1041–7.
38. Wang YB, Cui YX, Song JN, Yang Q, Wang G. Efficacies of various surgical regimens in the treatment of renal calculi patients: a network meta-analysis in 25 enrolled controlled clinical trials. Kidney Blood Press Res **2018**; 43:1183–94.
39. Brain E, Geraghty RM, Lovegrove CE, Yang B, Somani BK. Natural history of post-treatment kidney stone fragments: a systematic review and meta-analysis. J Urol **2021**; 206:526–38.
40. Bichler KH, Eipper E, Naber K, et al. Urinary infection stones. Int J Antimicrob Agents **2002**; 19:488–98.
41. Xu LHR, Adams-Huet B, Poindexter JR, et al. Temporal changes in kidney stone composition and in risk factors predisposing to stone formation. J Urol **2017**; 197: 1465–71.
42. Taylor EN, Stampfer MJ, Curhan GC. Dietary factors and the risk of incident kidney stones in men: new insights after 14 years of follow-up. J Am Soc Nephrol **2004**; 15:3225–32.
43. Vanguri RS, Luo J, Aukerman AT, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. Nat Cancer **2022**; 3:1151–64.
44. Coutant C, Olivier C, Lambaudie E, et al. Comparison of models to predict non-sentinel lymph node status in breast cancer patients with metastatic sentinel lymph nodes: a prospective multicenter study. J Clin Oncol **2010**; 21:2800–8.