Dovepress
Taylor & Francis Group

ORIGINAL RESEARCH

CECT-Based Radiomic Nomogram of Different Machine Learning Models for Differentiating Malignant and Benign Solid-Containing Renal Masses

Lu Qian [1],*, BinHai Fu [2],*, Hong He [2], Shan Liu [1], RenCai Lu [2]

[1]Department of Pathology, the First People's Hospital of Yunnan Province, The Affiliated Hospital of Kunming University of Science and Technology, Kunming, Yunnan, 650032, People's Republic of China; [2]Department of Nuclear Medicine, The First People's Hospital of Yunnan Province, the Affiliated Hospital of Kunming University of Science and Technology, Kunming, Yunnan, 650032, People's Republic of China

*These authors contributed equally to this work

Correspondence: RenCai Lu, Department of Nuclear Medicine, The First People's Hospital of Yunnan Province, the Affiliated Hospital of Kunming University of Science and Technology, No. 157 Jinbi Road, Kunming, Yunnan, 650032, People's Republic of China, Tel +86 087163614503, Email lrc_09@126.com

Objective: This study aimed to explore the value of a radiomic nomogram based on contrast-enhanced computed tomography (CECT) for differentiating benign and malignant solid-containing renal masses.

Materials and Methods: A total of 122 patients with pathologically confirmed benign (n=47) or malignant (n=75) solid-containing renal masses were enrolled in this study. Radiomic features were extracted from the arterial, venous and delayed phases and further analysed by dimensionality reduction and selection. Four mainstream machine learning algorithm training models, namely, support vector machine (SVM), k-nearest neighbour (kNN), light gradient boosting (LightGBM) and logistic regression (LR), were constructed to determine the best classifier model. Univariate and multivariate analyses were used to determine the best clinical characteristics for constructing a clinical model. The radiomic and clinical signatures were integrated to construct a combined radiomic nomogram model. Receiver operating characteristic (ROC) curves and the area under the curve (AUC) were used to evaluate the performance of the radiomic nomogram, radiomic signature, and clinical model.

Results: Thirteen radiomic features were selected for the development of the radiomic signature. Among the various radiomic models, the LR model demonstrated superior predictive efficiency and robustness, yielding an AUC of 0.952 in the training cohort and 0.887 in the test cohort. The AUC for the clinical model was 0.854 in the training cohort and 0.747 in the test cohort. Furthermore, the radiomic nomogram, which incorporated sex, age, alcohol consumption history, and the radiomic signature, exhibited excellent discriminative performance, yielding an AUC of 0.973 in the training cohort and 0.900 in the test cohort.

Conclusion: The radiomic nomogram based on CECT offers a promising and noninvasive approach for distinguishing malignant from benign solid renal masses. This tool can be used to guide treatment strategies effectively and can provide valuable insights for clinicians.

Keywords: computed tomography, renal neoplasm, radiomics, machine learning

Introduction

Renal masses (RMs) are biologically heterogenous and can be categorized as benign or malignant; based on their likelihood of aggressive malignancy, active surveillance and surgery are optional management methods.[1] Renal cell carcinoma (RCC) is common, and a surgical approach is often necessary; however, benign tumours or those with low malignant potential may not require treatment.[2–4] Preoperatively misclassified renal masses can lead to unnecessary surgery, and data suggest that there is a rapidly increasing trend of overtreatment.[5] Refraining from unnecessary surgical interventions transcends mere considerations of cost-effectiveness from the perspective of patients.[6]

Accurately differentiating malignant from benign RMs before surgery is important for follow-up management. Imaging tools such as ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) are widely used in the diagnosis of RMs.[7–9] Although each modality has advantages and characteristics, distinguishing malignant from benign RMs via conventional imaging methods[4] is still challenging. Nuclear medicine, such as single photon emission computed tomography/computed tomography (SPECT/CT) and positron emission tomography computed tomography (PET-CT) molecular imaging, showed great value in evaluating renal lesions, but there are still some drawbacks to a generalized use.[10] Transmembrane protein mucin-1 (MUC1) might be utilized in clinical settings as a diagnostic marker for renal malignancy,[11] however, the guidance of MUC1 for clinical is still deficient. In particular, solid-containing RMs such as oncocytomas and lipid-poor angiomyolipomas (AMLs) can be differentiated from RCC.[4,6] Among the medical imaging modalities, contrast-enhanced CT (CECT) is currently the most commonly used method for evaluating RMs.[4]

Artificial intelligence (AI) has become a key technology for analysing multi-parametric data, with applications in medical imaging, including image acquisition, post-processing, and data mining and modelling.[12] Radiomics is a promising method that involves the development of algorithms and AI and provides high-throughput extraction of quantitative features.[13] AI holds great potential to enhance cancer imaging interpretation, inference of tumor genotype and biology from radiographic phenotypes, outcome prediction, and evaluation of disease and treatment effects on adjacent organs. AI may also automate initial image interpretation and transform clinical workflows in radiographic detection, management decisions, and post-intervention monitoring.[14,15] Subsequent analysis of these features can facilitate clinical decisions, such as improving diagnostic, prognostic and predictive accuracy.[13] Machine learning represents a subset of artificial intelligence that employs algorithms to iteratively learn from and make predictions on data, thereby identifying complex patterns that may elude human cognition.[16] Researchers have employed deep learning or radiomics approaches to differentiate clear cell renal cell carcinoma (ccRCC) from other benign renal tumors.[17–19] Nevertheless, few studies have examined the use of different machine learning models to analyse CECT images and differentiate benign and malignant solids containing RMs.

Herein, different machine learning models were tested, and the best model was combined with a clinical model to explore its value in distinguishing benign and malignant solids containing RMs.

Methods

Patients

A total of 122 patients with renal neoplasms diagnosed between December 2019 and December 2022 were retrospectively enrolled in our study. The inclusion criteria for patients were as follows: (1) underwent abdominal three-phase CECT before surgery, including the arterial phase (20–25 s), venous phase (70–90 s) and delayed phase (180 s); (2) patients had no other treatment history, such as chemotherapy or radiotherapy, before CECT; (3) the postoperative pathology diagnosis and clinical data were complete; (4) solid components were found in the renal tumours. The exclusion criteria were as follows: (1) incomplete three-phase CECT; (2) patients without any solid components in renal tumours; (3) patients whose pathology and clinical data were incomplete. A total of 75 patients with malignant solid-containing RMs were included in the study, including: Renal clear cell carcinoma (n=58, 77.33%), urothelial carcinoma (n=3, 4.00%), chromophobe renal cell carcinoma (n=6, 8.00%), leiomyosarcoma (n=2, 2.67%), papillary renal cell carcinoma (n=3, 4.00%), multilocular cystic renal cell carcinoma (n=1, 1.33%), oncocytic papillary renal cell carcinoma (n=1, 1.33%), sarcomatous renal cell carcinoma (n=1, 1.33%). A total of 47 patients with benign solid-containing RMs were included in the study, including renal angiomyolipoma (n=40, 74.47%), renal oncocytoma (n=7, 14.89%), perivascular epithelioid cell tumor (n=3, 6.38%), complex renal cyst (n=2, 4.26%). The patients were randomly allocated to a training cohort and a test cohort at a 7:3 ratio, and the flowchart of patient selection is shown in Figure 1.

This study was conducted in accordance with the declaration of Helsinki and approved by the Ethics Committee of the First People's Hospital of Yunnan Province (No. KHLL2024-KY008). Due to the study's methodology was retrospective analysis, which did not entail direct patient involvement, the Ethics Committee of the First People's Hospital of Yunnan Province, China, sanctioned the study's protocol and exempted the requirement for acquiring informed consent from participants. All patients' data was anonymized or maintained with confidentiality.
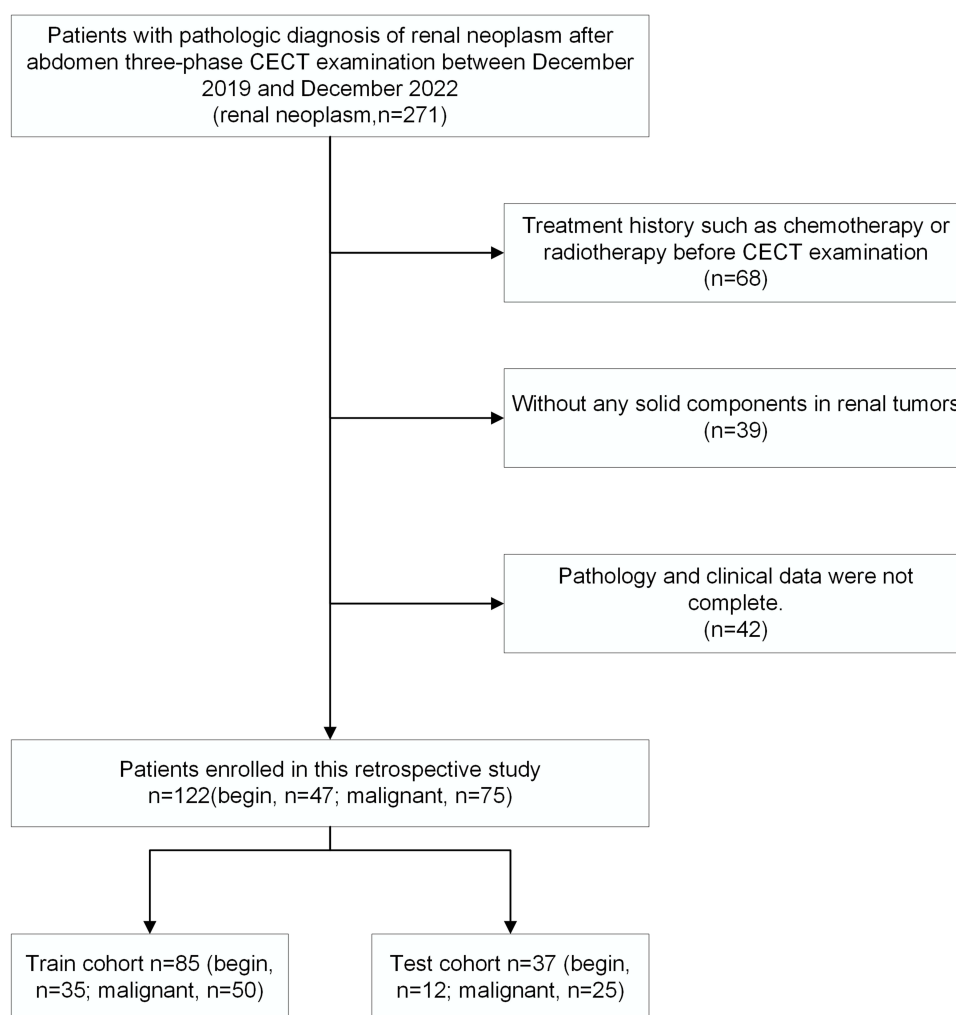
Patients with pathologic diagnosis of renal neoplasm after abdomen three-phase CECT examination between December 2019 and December 2022 (renal neoplasm,n=271)

Treatment history such as chemotherapy or radiotherapy before CECT examination (n=68)

Without any solid components in renal tumors (n=39)

Pathology and clinical data were not complete. (n=42)

Patients enrolled in this retrospective study n=122(begin, n=47; malignant, n=75)

Train cohort n=85 (begin, n=35; malignant, n=50)

Test cohort n=37 (begin, n=12; malignant, n=25)

Figure 1 Flowchart for patient selection in our study.

Image Data Acquisition and Clinical Data Collection

Siemens drive CT scanners were used, and the scanning parameters were as follows: tube voltage, 120 kV; tube current, 150–200 mAs; matrix, 512 × 512; section thickness, 1.5 mm; and section interval, 3 mm. The scan area included at least the entire kidneys. After the nonenhanced CT scan, three-phase CECT was performed by injecting nonionic iodinated contrast medium (iohexol, 350 mg/I) at a dose of 1.5 mL/kg and a speed of 3.5 mL/s. CECT images were acquired in the arterial phase (20–25 s), venous phase (70–90 s) and delayed phase (180 s). All DICOM format images were transferred to the Neusoft picture archiving and communication systems (PACSs). Clinical data, including sex, age, smoking history, drinking history, history of hypertension, history of diabetes and tumour shape, were collected simultaneously. Two radiologists with 8 and 6 years of experience in abdominal CECT were assessed, and the images were agreed upon. The largest lesion with confirmed pathology was selected for subsequent analysis if multiple lesions were present.

Radiomics Analysis of CECT Images
Tumour Segmentation and Feature Extraction

A typical workflow of our study is shown in Figure 2. All DICOM-formatted CT images were saved as a window of 400 hU and a window of 40 hU. Two radiologists segmented the volume of interest (VOI) manually in the arterial phase, venous phase and delayed phase; they were all blinded to the patients' histopathological results before segmentation. 3D Slicer software (version 5.3.0, https://www.slicer.org/) was used, and the tumours were delineated along margins layer-by
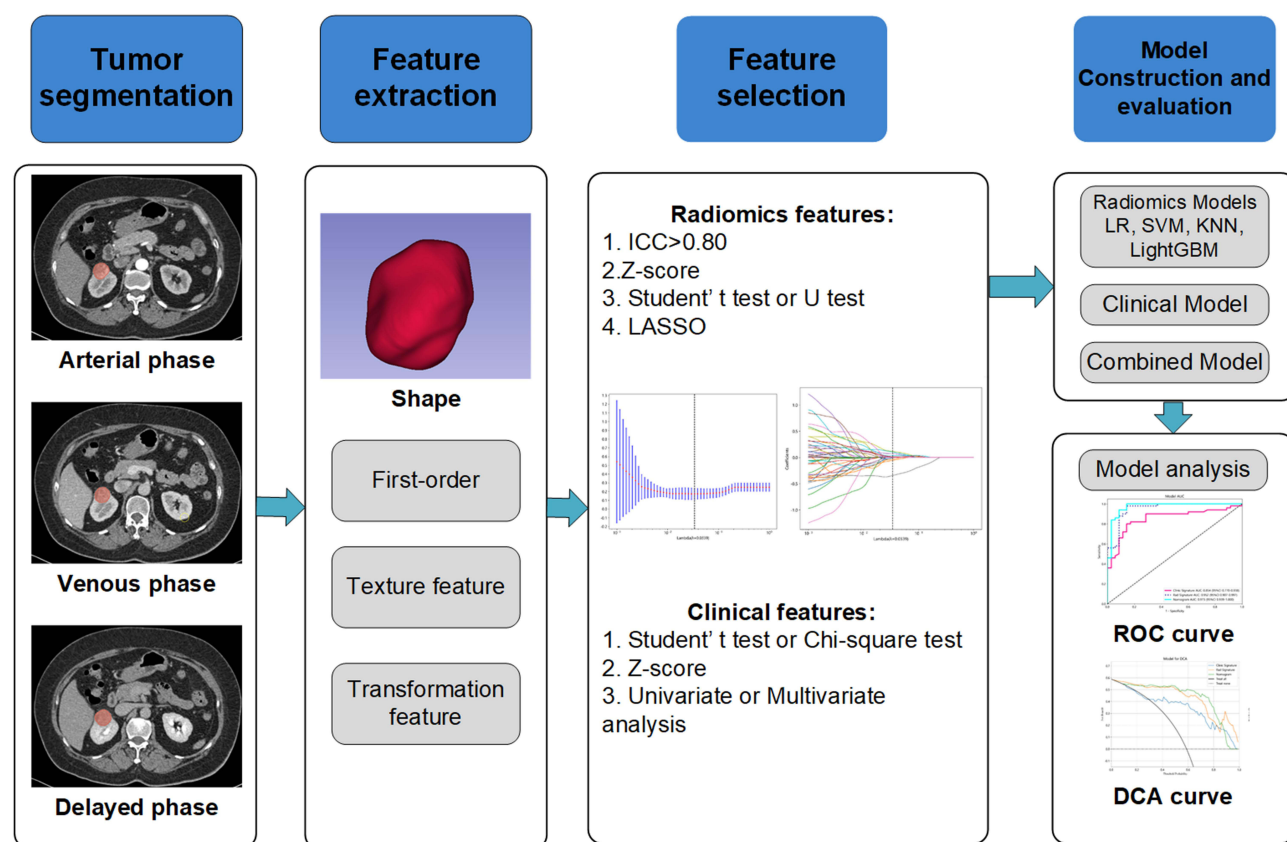
Figure 2 Workflow of the key steps in our study.

-layer on cross-sectional CT images. VOIs were saved in NIFIT format for further analysis. Features were extracted by using PyRadiomics in Python (https://pyradiomics.readthedocs.io/en/2.1.2/).

In our study, we classified handcrafted features into three main categories: geometry, intensity, and texture, each targeting distinct aspects of tumor characteristics. Geometric features provide insights into the tumor's shape and structure, revealing its morphological properties. Intensity features examine the brightness levels of voxels within the tumor, offering information on the distribution of intensities. Texture features, derived using methods such as the Gray Level Co-occurrence Matrix (GLCM), grey-level dependence matrix (GLDM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM), capture the spatial patterns within the tumor. Together, these feature categories enable a comprehensive analysis of tumor characteristics, enhancing our understanding of its underlying properties. Finally, the following radiomic features were extracted: shape (n = 42), first order (n = 54), GLCM (n = 72), GLDM (n = 42), GLRLM (n = 48), GLSZM (n = 48), and NGTDM (n = 15). All radiomic features and their corresponding significance were derived in accordance with the guidelines of the Image Biomarkers Standardization Initiative (IBSI).[20]

In order to evaluate the reproducibility of radiomic feature extraction, both intra- and interobserver consistency were examined using intra- and interclass correlation coefficients (ICC). A set of 40 images was randomly selected for region of interest (ROI) segmentation, performed by two radiologists, each with more than six years of experience in abdominal image interpretation. The interobserver reliability was evaluated by comparing the results of radiologists A and B, while the interobserver consistency was assessed by having radiologist A repeat the segmentation process. Intra- and interclass correlation coefficient (ICC) values > 0.8 was regarded as demonstrating strong agreement in feature extraction.

Radiomics Feature Selection Signature Model Development

Z-score normalization was used to reduce the bias of radiomic features. Features were selected by the following steps: (1) features with an ICC > 0.8 were identified; (2) Student's *t* test or *U*-test was used to compare the extracted features of two groups, and features with P values < 0.05 were maintained for further analysis; and (3) Pearson's correction coefficients were calculated to identify redundant and collinear features, and features with mutual correlation coefficients > 0.9 were excluded; and (4) the least absolute shrinkage and selection operator (LASSO) regression model with 10-fold cross-validation was used to select optimal features. LASSO regression is a powerful technique for variable selection that effectively reduces the dimensionality of the model by eliminating irrelevant predictors, while preserving model accuracy. This not only improves the interpretability of the model but also enhances its predictive performance.[21]

Radiomic models were built from the final selected features. Four mainstream machine learning algorithm training models were used to construct the best classifier model: support vector machine (SVM), k-nearest neighbour (kNN), light gradient boosting (LightGBM) and logistic regression (LR). SVM is a supervised machine learning model grounded in the Vapnik–Chervonenkis (VC) dimension, and it is proficient in making robust predictions.[22] kNN is a classification method that relies on the proximity to nearby data points. The fundamental concept is rooted in statistical theory, where a point is classified by evaluating the weights of its neighbors and assigning it to the category with the highest weight.[23] LightGBM is a fast, scalable, and high-performance gradient boosting framework designed for distributed environments. It offers quick training, high accuracy, and excellent scalability, making it ideal for handling large-scale, high-dimensional datasets in tasks like regression, classification, and ranking.[24] LR model predicts the likelihood of a binary dependent variable based on other variables.[25] The area under the receiver operating characteristic (ROC) curve (AUC) was calculated to compare the diagnostic performance of the two methods. The sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) were also calculated. The best radiomic model was then demonstrated.

Clinical and Combined Model Construction and Validation

Univariate analysis was used to determine the optimal clinical characteristics for differentiating between benign and malignant lesions in the training cohort. Parameters with a P value < 0.05 were chosen for the construction of clinical models. Odds ratios (ORs) with 95% confidence intervals (CIs) were calculated simultaneously for each characteristic. The best radiomic models provided by machine learning were used to construct a combined model with a clinical model. The diagnostic performance of the models was assessed with ROCs, including the area under the curve (AUC), accuracy, sensitivity, specificity, PPV, and NPV. Decision curve analysis (DCA) was used in the training and test cohorts to evaluate the models' ability to predict tumour differentiation.

Statistical Analysis

All the statistical analyses were performed with Python software (version 3.4; http://www.python.org). Continuous variables are expressed as $\bar{x} \pm s$ and were compared by Student's *t* test. Categorical variables are expressed as numbers (n%) and were compared by the chi-square test or Fisher's exact test. A two-sided *P* value < 0.05 was considered to indicate statistical significance.

Results

Clinical Data

A total of 122 patients were enrolled in our study. The ratio of malignant renal masses in the training cohort and test cohort was not significantly different (50/85 vs 25/37, *P*=0.362). Patients were randomly divided into a training cohort (n=85) and a test cohort (n=37). A comparison of the clinical data between the two groups is shown in Table 1. All the clinical data were significantly different except for shape in the training cohort (*P*<0.05). Only age was significantly different in the test cohort (*P*<0.05). Univariate and multivariate analyses showed that sex, age and alcohol consumption were independent predictors of malignant masses (Table 2). These three features were further used to construct the clinical model.

Table I Clinical Characteristics of the Patients in the Training and Test Cohorts

Characteristics	Training Cohort (n=85)			Test Cohort (n=37)		P value
	Benign (n=35)	Malignant (n=50)		Benign (n=12)	Malignant (n=25)	
Age[#]	46.49±11.59	54.24±14.26	0.009*	45.42±13.24	54.36±11.86	0.046*
Gender			<0.001*			0.162
Female	26(74.29)	17(34.00)		8(66.67)	9(36.00)	
Male	9(25.71)	33(66.00)		4(33.33)	16(64.00)	
Smoking			0.049*			0.769
No	30(85.71)	32(64.00)		9(75.00)	16(64.00)	
Yes	5(14.29)	18(36.00)		3(25.00)	9(36.00)	
Drinking			0.021*			0.249
No	33(94.29)	36(72.00)		12(100.00)	20(80.00)	
Yes	2(5.71)	14(28.00)		0	5(20.00)	
Hypertension			<0.001*			1.000
No	32(91.43)	26(52.00)		9(75.00)	18(72.00)	
Yes	3(8.57)	24(48.00)		3(25.00)	7(28.00)	
Diabetes			0.047*			1.000
No	34(97.14)	40(80.00)		12(100.00)	25(100.00)	
Yes	1(2.86)	10(20.00)		0	0	
Shape			0.736			1.000
Regular	21(60.00)	33(66.00)		8(66.67)	16(64.00)	
Irregular	14(40.00)	17(34.00)		4(33.33)	9(36.00)	

Notes: Unless otherwise indicated, the data are presented as numbers (%). Student's t test was used for continuous data. Categorical data were compared by using the χ^2 test or Fisher's exact test, as possible. [#]Data are means ± standard deviations. *Data are statistically significant.

Table 2 Univariate and Multivariate Logistic Regression Analyses for Selecting Clinical Features for Model Construction

Characteristics	Univariate Analysis		Multivariate Analysis	
	OR (95% CI)	P value	OR (95% CI)	P value
Gender	1.429(1.246, 1.639)	0.000*	1.405(1.196, 1.65)	0.001*
Age	1.011(1.005, 1.016)	0.001*	1.009(1.004, 1.014)	0.006*
Smoking	1.246(1.062, 1.461)	0.024*	0.877(0.73, 1.055)	0.242
Drinking	1.419(1.177, 1.713)	0.002*	1.278(1.051, 1.553)	0.039*
Hypertension	1.377(1.182, 1.605)	0.001*	1.205(1.029, 1.411)	0.052
Diabetes	1.382(1.074, 1.779)	0.036*	1.095(0.862, 1.391)	0.529
Shape	0.963(0.826, 1.123)	0.687		

Note: *Data are statistically significant.
Abbreviations: OR, odds ratio; CI, confidence interval.

Radiomic and Signature Models and Performance

A total of 322 features were extracted from the VOI in three phases. Thirteen features with nonzero coefficients were retained after feature downscaling and selection (Figure 3). The ROC curves of the four mainstream machine learning algorithm training models (LR, SVM, KNN and LightGBM) in the training cohort and test cohort are shown in Table 3 and Figure 4. The best radiomic model in the training cohort was SVM (AUC, 0.953 [95% CI, 0.8981–1.0000]; accuracy, 0.906; sensitivity, 0.900; specificity, 0.914; PPV, 0.937; NPV, 0.865).

Among the various models evaluated, the logistic regression (LR) model exhibited superior performance across all cohorts, with an AUC of 0.952 (95% CI, 0.9072–0.9968) for the training cohort and 0.887 (95% CI, 0.7782–0.9951) for the test cohort (accuracy, 0.784; sensitivity, 0.720; specificity, 0.917; PPV, 0.947; NPV, 0.611). This consistent high
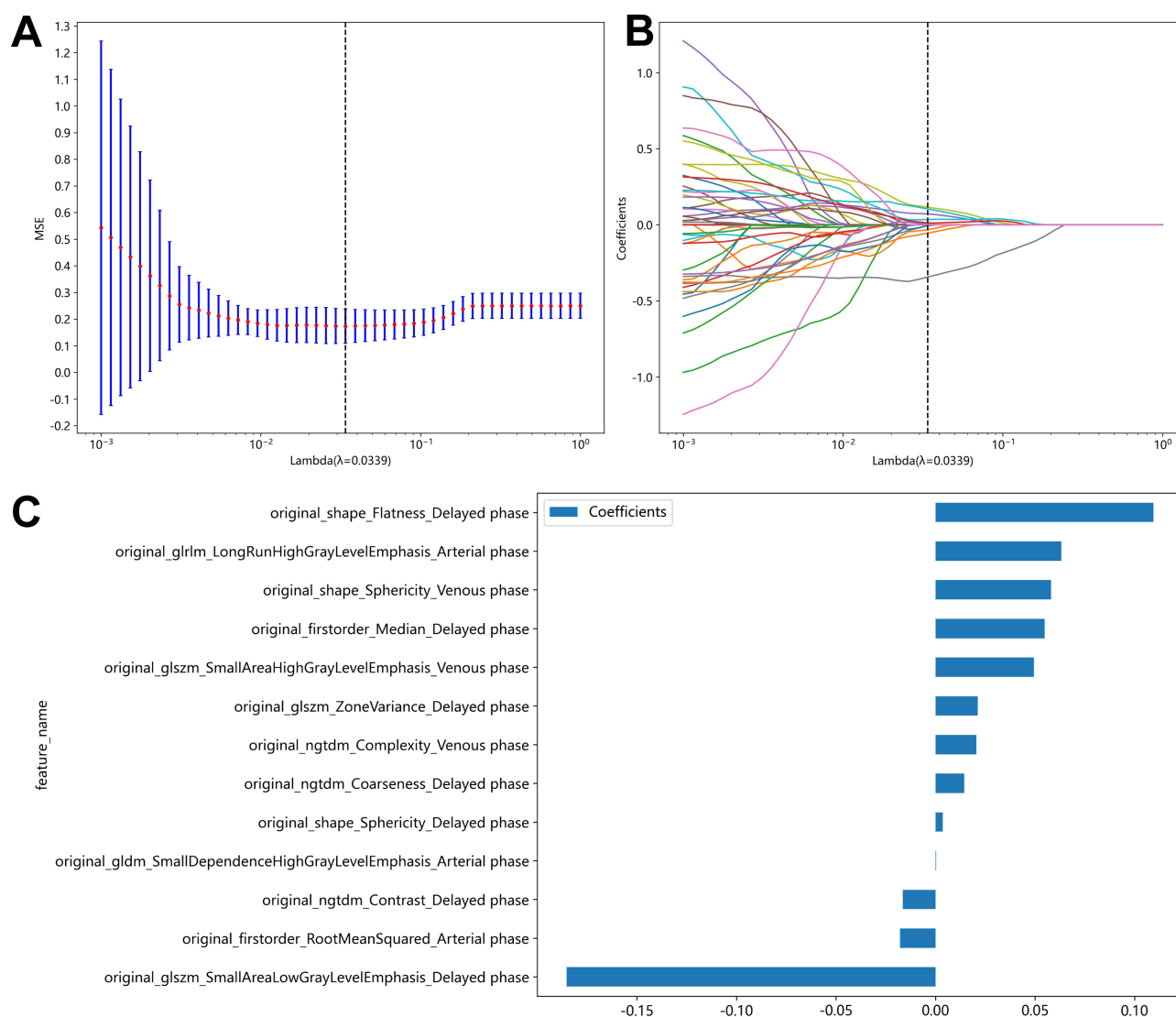
Figure 3 Feature selection with the least absolute shrinkage and selection operator (LASSO) regression model. (**A**) The penalization parameter λ was selected with 10-fold cross-validation as the minimum criterion in the LASSO model. The vertical lines indicate the optimal value of the LASSO tuning parameter (λ=0.0339). (**B**) LASSO coefficient profile of the radiomic features. The vertical dashed lines indicate 13 radiomic features with nonzero coefficients. (**C**) The 13 features contributing to radiomic features based on CECT images weighted by standardized regression according to the LASSO model were used to differentiate malignant from benign RMs.

performance across different data subsets underscores the robustness and efficacy of the LR model in differentiating between the two classes in our study. Furthermore, the LR model's ability to maintain a relatively high AUC in the test cohort, compared to other models, suggests its strong potential for generalizing to unseen data. This makes it a promising tool for similar binary classification tasks in future research. Based on these considerations, we selected the LR model for further analysis to ensure model stability.

Combined Model Construction and Performance

In the analysis of fused models combining clinical and radiomic modalities, we observe that the radiomics nomogram model demonstrates better performance in the test cohort compared to other models. The diagnostic performance and ROC curves of the three models are shown in Table 3 and Figure 5. The AUC improved to 0.973 (95% CI, 0.9388–1.0000) in the training cohort and 0.900 (95% CI, 0.7874–1.0000) in the test cohort (accuracy, 0.865; sensitivity, 0.880; specificity, 0.833; PPV, 0.917; NPV, 0.769), which was better than that of the clinical or radiomic signature models. The DCA and calibration curves for both the training and test cohorts of the three models are shown in Figure 6.

Table 3 Diagnostic Performance of Different Models for Differentiating Benign and Malignant Solids Containing Renal Masses in the Training and Test Cohorts

Model	Group	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV
LR	Training	0.929	0.952	0.9072–0.9968	0.980	0.857	0.907	0.968
	Test	0.784	0.887	0.7782–0.9951	0.720	0.917	0.947	0.611
SVM	Training	0.906	0.953	0.8981–1.0000	0.900	0.914	0.937	0.865
	Test	0.811	0.830	0.6882–0.9718	0.880	0.667	0.846	0.727
KNN	Training	0.847	0.941	0.8975–0.9837	0.800	0.914	0.93	0.762
	Test	0.757	0.883	0.7819–0.9848	0.640	1	1	0.571
LightGBM	Training	0.871	0.915	0.8487–0.9811	0.920	0.824	0.868	0.875
	Test	0.838	0.872	0.7577–0.9856	0.880	0.75	0.88	0.75
Clinic model	Training	0.824	0.854	0.7696–0.9378	0.800	0.857	0.889	0.75
	Test	0.784	0.747	0.5660–0.9274	0.800	0.75	0.87	0.643
Radiomics Nomogram*	Training	0.941	0.973	0.9388–1.0000	1.000	0.857	0.909	1.000
	Test	0.865	0.900	0.7874–1.0000	0.880	0.833	0.917	0.769

Note: *Represents models were constructed using LR.
Abbreviations: LR, logistic regression; SVM, support vector machine; KNN, k nearest neighbour; LightGBM, light gradient boosting; AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

The analysis of these curves indicates that our Combined model offers a notable advantage in terms of net benefit derived from predicted probabilities.

Discussion

Solid-containing RMs, including partial fat-containing and completely solid RMs, were ultimately selected for this study. RMs with complete fat composition were not included in our study because they are easy to detect and distinguish via conventional imaging methods.[26] However, some malignant tumours, such as liposarcoma, Wilms tumours and renal cell carcinoma (RCC), may contain fat, which leads to challenges in preoperative diagnosis.[26] Additionally, AMLs with poor fat weight, epithelial cysts or haemorrhagic complications have different appearances, and their imaging manifestations may overlap with those of malignant renal tumours.[27]

Xu Wang[9] developed a prediction model using thin-section multidetector CT to distinguish hypervascular ultrasmall renal cell carcinoma (≤2 cm) from renal angiomyolipoma with minimal fat in the early stage. However, their study did not encompass renal cell carcinoma (RCC) larger than 2 cm or other hypervascular renal tumours, such as renal oncocytoma (RO). Due to the shared origin of some benign and malignant renal tumours, they exhibit overlapping histologic and imaging characteristics.[28] Consequently, distinguishing benign renal tumours from malignant renal tumours using conventional enhanced imaging techniques is still challenging.

Radiomics emphasizes that medical images are not only visual representations but also valuable sources of data that can be analysed and utilized for improving patient care and outcomes through a deeper understanding of the underlying data present.[13] Xue-Ying Sun[29] compared the performance of machine learning models with that of expert radiologists, and their results revealed that radiologic-radiomic machine learning could improve interreader concordance and performance, with a high sensitivity of 86.3% and 73.4% and a specificity of 83.3% and 91.7%, respectively, for differentiating clear cell RCC and chromophobe RCC from fat-poor AML and RO. C. Erdim[30] conducted an analysis of texture patterns in CECT images using machine learning algorithms to differentiate malignant RMs from benign RMs. Their results showed a high accuracy of 90.5% in distinguishing them via the random forest (RF) algorithm. However, they also noted that two sets of patients (one for training and one for testing) are needed for further analysis.[30] A recent study also showed that enhanced diagnostic accuracy and efficiency are achieved through the utilization of CT texture-based machine learning algorithms to differentiate between benign and malignant cystic RMs.[31] A study by Cassandre Garnier[32] also demonstrated the significant value of radiomics models based on CECT in predicting malignant and benign renal tumours. All these studies support the potential of machine learning algorithms to help radiologists to achieve more accurate diagnostic outcomes.
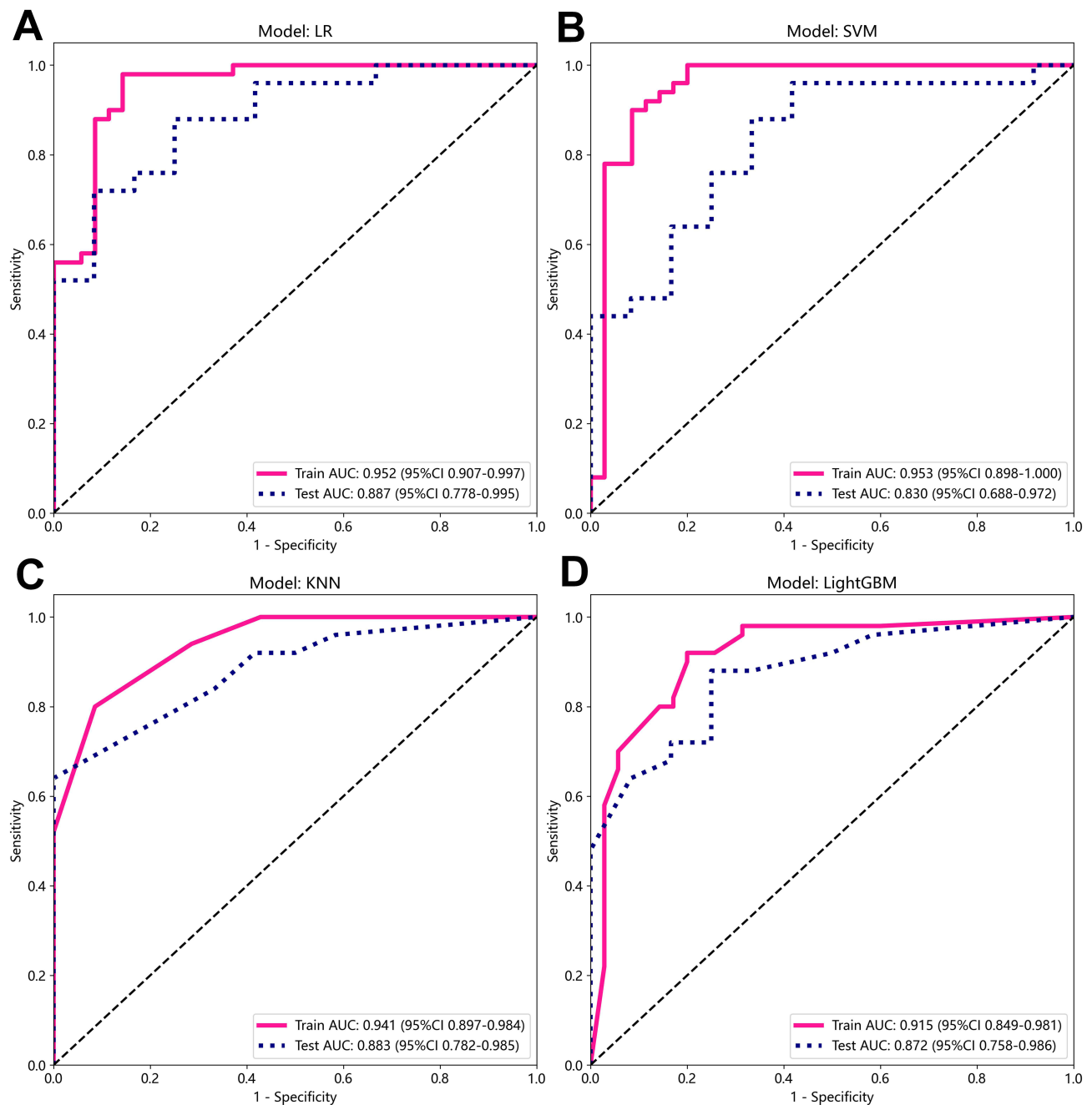
Figure 4 The receiver operating characteristic (ROC) curves of the LR (**A**), SVM (**B**), KNN (**C**), and LightGBM (**D**) models in the training and test cohorts. The SVM model (**B**) achieved the highest AUC in the training cohort (0.953, 95% CI: 0.898–1.000), while the LR model (**A**) demonstrated the highest AUC in the test cohort (0.887, 95% CI: 0.778–0.995).

Our study indicated that CECT-based machine learning models demonstrate potential value in differentiating benign and malignant solids containing RMs. In our study, we developed a radiomic signature comprising 13 radiomic features that exhibited satisfactory predictive ability for distinguishing malignant from benign RM. Among these factors, the original shape flatness (OSF) achieves high predictive performance and holds high variable importance in the radiomic model for distinguishing malignant from benign RM. This result is similar to the findings reported by Yong Huang,[33] where OSF constituted the primary factor for the construction of machine learning models to develop personalized treatment decision models for patients with inoperable oesophageal cancer. The flatness of an object was quantified by OSF, where higher values indicate a greater degree of flatness.[33] The grey-level size-zone matrix (GLSZM) and small area low grey-level emphasis (SALGLE) exhibited the most negative correlations (Figure 3) in the predictive model. Yusufaly[34] also reported that the SALGLE score was strongly negatively
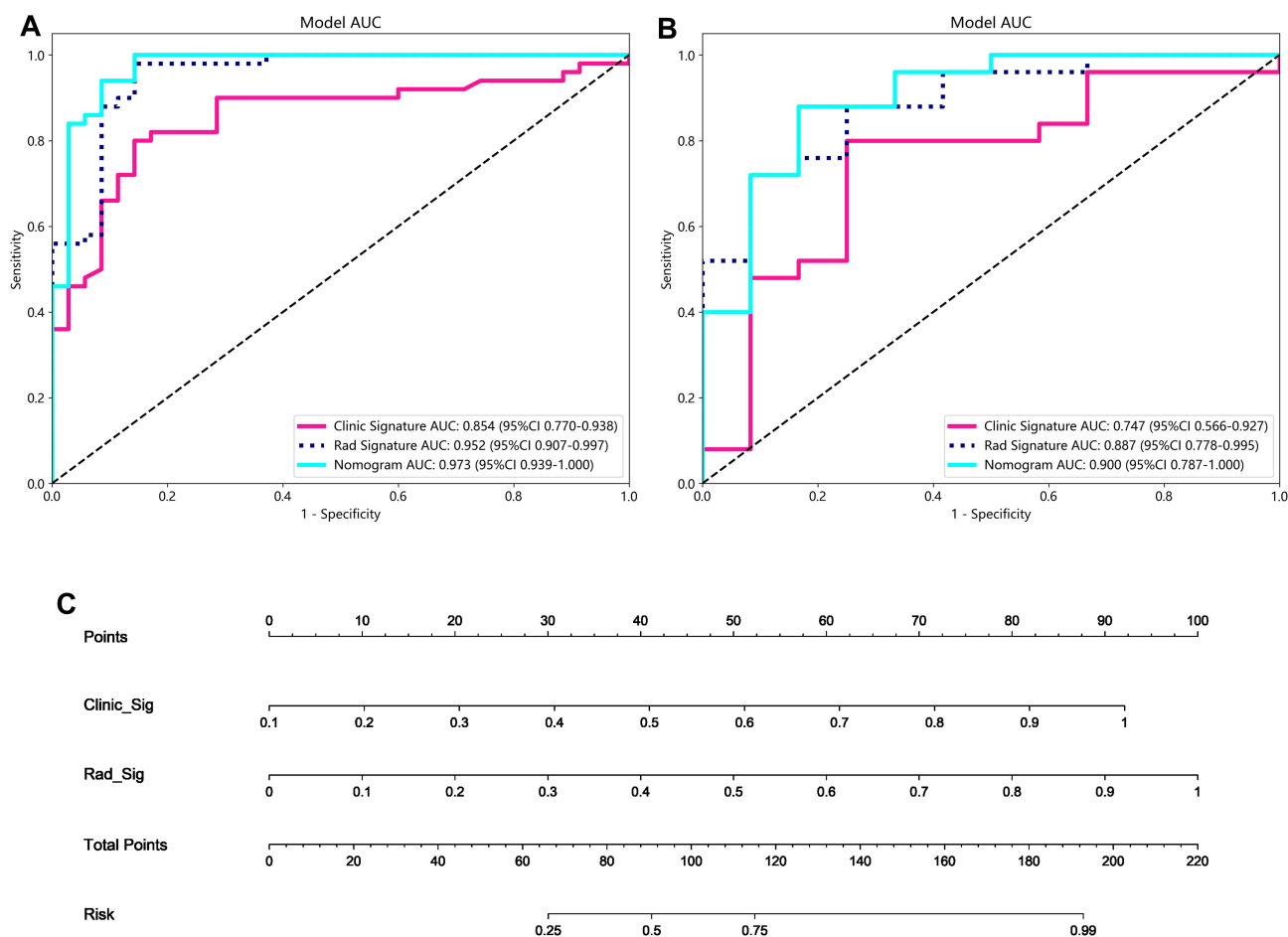
Figure 5 Receiver operating characteristic (ROC) curves for the clinical signature model, radiomic signature model, and radiomic nomogram model for predicting malignancy in the training cohort (**A**) and test cohort (**B**). The radiomics nomogram incorporating the clinical signature and radiomics signature (**C**).

correlated with the outcome of treatment in patients with previously untreated locoregionally advanced cervical cancer according to a radiomic model. Other features, including the GLDM, NGTDM, GLRLM and first-order features, have also been used to construct radiomic models. Overall, the LR model achieved the best performance, with an AUC of 0.952 in the training cohort and 0.887 in the test cohort (Table 3, Figure 4).

Univariate and multivariate analyses revealed that sex, age, and alcohol consumption were significant predictors for distinguishing between malignant and benign RMs (Table 2). These predictors have the potential to establish both a clinical model and a combined model for identification purposes. The combined model of the radiomic signature and clinical features achieved better discrimination. This finding is consistent with the findings of N. Nassiri,[35] who constructed a CT-based radiomic model and achieved good results in distinguishing between malignant and benign RMs. However, they used only the random forest (RF) and REAL AdaBoost platforms for model construction, and the REAL AdaBoost model achieved better diagnostic performance (AUC of 0.83),[35] which is different from our research findings where the LR model was optimal. The incorporation of clinical factors into radiomic analysis led to improved performance, with an AUC of 0.973 (95% CI, 0.9388–1.0000) in the training cohort and 0.900 (95% CI, 0.7874–1.0000) in the test cohort for the combined features (Table 3, Figure 5). The above results suggest that the combination of clinical models and machine learning models can effectively predict malignant and benign solids containing RMs.

Our study demonstrates that the nomogram constructed based on clinical and radiomic features has good applicability in predicting solid-containing RMs, aiding clinicians in making more accurate decisions in practice and providing personalized risk assessments for patients, thus avoiding unnecessary overtreatment. However, it is important to acknowledge that different studies may employ varying enhancement scanning protocols and image segmentation methods, which could potentially limit the clinical applicability of the study's findings.[36] Furthermore, there remains an opportunity for further integration of radiomic data with
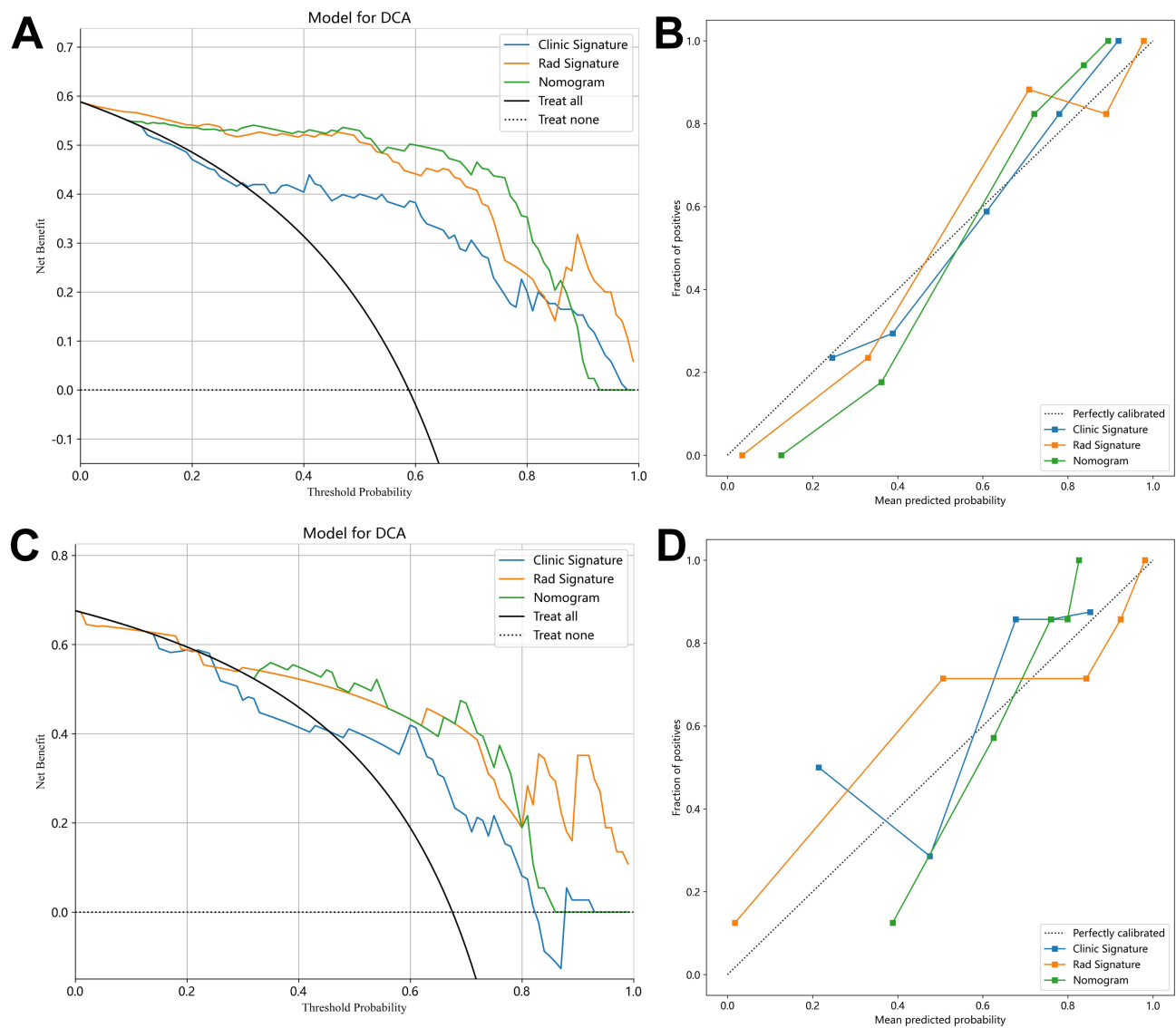
Figure 6 Decision curve analysis (DCA) and calibration curves for three models for differentiating benign and malignant solids containing renal masses in the training (**A** and **B**) and test cohorts (**C** and **D**). The graphs show that the combined nomogram has the greatest net benefit for both datasets.

clinical, genetic, and metabolic data to gain a more comprehensive understanding of renal masses, and to leverage the complementary value of each modality in the diagnosis, prognosis, prediction of treatment response, and monitoring of renal masses.[37]

This study has several limitations that should be acknowledged. First, this was a retrospective, single-centre study, the relatively small sample size may introduce potential selection bias. Second, the random forest (RF) model was not used in our study because RF has poor generalizability when processing small sample cohorts and may overfit.[38,39] Third, this study did not utilize non-CECT images for analysis, as many RMs cannot be distinguished from normal renal parenchyma on noncontrast scans; however, this does not disregard the value of noncontrast scans in the diagnosis of renal tumours. Additionally, we did not utilize a separate validation set. Instead, we employed ten-fold cross-validation to assess the model's performance, ensuring that the model was tested on different subsets of the data, thereby reducing the risk of overfitting. The independent test set was used solely for the final performance evaluation to simulate the model's ability to generalize to unseen data. Thus, an external validation is essential in future research to improve the reliability of the research findings. These factors may all serve to limit the generalizability of the results. It is necessary for future studies to be conducted with larger sample sizes and across multiple centers to enhance the reliability and determine the efficacy of machine learning models in distinguishing between malignant and benign solid-containing RMs.

Conclusion

In summary, this study introduces machine learning models that integrate both radiomic signatures and clinical factors, thereby offering a noninvasive and promising approach for the preoperative differentiation of malignant tissue from benign solid tissue containing RM. These enhancements may refine the accuracy of clinical decision-making and potentially avert the risks associated with unwarranted aggressive surgical treatments. However, it is crucial to highlight the imperative need for external validation and multicenter studies to substantially enhance the reliability of the research findings.

Ethics Statement

This study was conducted in accordance with the declaration of Helsinki and approved by the Ethics Committee of the First People's Hospital of Yunnan Province (No. KHLL2024-KY008). Due to the study's methodology was retrospective analysis, which did not entail direct patient involvement, the Ethics Committee of the First People's Hospital of Yunnan Province, China, sanctioned the study's protocol and exempted the requirement for acquiring informed consent from participants. All patients' data was anonymized or maintained with confidentiality.

Disclosure

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Pierorazio PM, Johnson MH, Patel HD, et al. Management of renal masses and localized renal cancer: systematic review and meta-analysis. *J Urol*. 2016;196(4):989–999. doi:10.1016/j.juro.2016.04.081
2. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca a Cancer J Clinicians*. 2021;71(3):209–249. doi:10.3322/caac.21660
3. Perazella MA, Dreicer R, Rosner MH. Renal cell carcinoma for the nephrologist. *Kidney Int*. 2018;94(3):471–483. doi:10.1016/j.kint.2018.01.023
4. Roussel E, Capitanio U, Kutikov A, et al. Novel imaging methods for renal mass characterization: a collaborative review. *Europ Urol*. 2022;81 (5):476–488. doi:10.1016/j.eururo.2022.01.040
5. Johnson DC, Vukina J, Smith AB, et al. Preoperatively misclassified, surgically removed benign renal masses: a systematic review of surgical series and United States population level burden estimate. *J Urol*. 2015;193(1):30–35. doi:10.1016/j.juro.2014.07.102
6. Kim JH, Li S, Khandwala Y, Chung KJ, Park HK, Chung BI. Association of prevalence of benign pathologic findings after partial nephrectomy with preoperative imaging patterns in the United States from 2007 to 2014. *JAMA Surg*. 2019;154(3):225. doi:10.1001/jamasurg.2018.4602
7. Zhu D, Li J, Li Y, et al. Multimodal ultrasound fusion network for differentiating between benign and malignant solid renal tumors. *Front Mol Biosci*. 2022;9. doi:10.3389/fmolb.2022.982703
8. Ding Y, Tan Q, Mao W, et al. Differentiating between malignant and benign renal tumors: do IVIM and diffusion kurtosis imaging perform better than DWI? *Eur Radiol*. 2019;29(12):6930–6939. doi:10.1007/s00330-019-06240-6
9. Wang X, Song G, Sun J, Shao G. Differential diagnosis of hypervascular ultra-small renal cell carcinoma and renal angiomyolipoma with minimal fat in early stage by using thin-section multidetector computed tomography. *Abdom Radiol*. 2020;45(11):3849–3859. doi:10.1007/s00261-020-02542-2
10. Tataru OS, Marchioni M, Crocetto F, et al. Molecular imaging diagnosis of renal cancer using (99m)Tc-Sestamibi SPECT/CT and Girentuximab PET-CT-current evidence and future development of novel techniques. *Diagnostics*. 2023;13(4). doi:10.3390/diagnostics13040593
11. Milella M, Rutigliano M, Lasorsa F, et al. The role of MUC1 in renal cell carcinoma. *Biomolecules*. 2024;14(3):315. doi:10.3390/biom14030315
12. Shiyam Sundar LK, Muzik O, Buvat I, Bidaut L, Beyer T. Potentials and caveats of AI in hybrid imaging. *Methods*. 2021;188:4–19. doi:10.1016/j.ymeth.2020.10.004
13. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563–577. doi:10.1148/radiol.2015151169

14. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. 2019;69 (2):127–157. doi:10.3322/caac.21552

15. Paudyal R, Shah AD, Akin O, et al. Artificial intelligence in CT and MR imaging for oncological applications. *Cancers*. 2023;15(9):2573. doi:10.3390/cancers15092573

16. Gaur K, Jagtap MM. Role of artificial intelligence and machine learning in prediction, diagnosis, and prognosis of cancer. *Cureus*. 2022. doi:10.7759/cureus.31008

17. Baghdadi A, Aldhaam NA, Elsayed AS, et al. Automated differentiation of benign renal oncocytoma and chromophobe renal cell carcinoma on computed tomography using deep learning. *BJU Int*. 2020;125(4):553–560. doi:10.1111/bju.14985

18. Schieda N, Nguyen K, Thornhill RE, McInnes MDF, Wu M, James N. Importance of phase enhancement for machine learning classification of solid renal masses using texture analysis features at multi-phasic CT. *Abdom Radiol*. 2020;45(9):2786–2796. doi:10.1007/s00261-020-02632-1

19. Tanaka T, Huang Y, Marukawa Y, et al. Differentiation of small (≤ 4 cm) renal masses on multiphase contrast-enhanced CT by deep learning. *Am J Roentgenol*. 2020;214(3):605–612. doi:10.2214/ajr.19.22074

20. Zwanenburg A, Vallieres M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328–338. doi:10.1148/radiol.2020191145

21. Li N, Li YL, Shao JM, Wang CH, Li SB, Jiang Y. Optimizing early neurological deterioration prediction in acute ischemic stroke patients following intravenous thrombolysis: a LASSO regression model approach. *Front Neurosci*. 2024;18:1390117. doi:10.3389/fnins.2024.1390117

22. Valkenborg D, Rousseau A-J, Geubbelmans M, Burzykowski T. Support vector machines. *Am J Orthodontics Dentofacial Orthopedics*. 2023;164 (5):754–757. doi:10.1016/j.ajodo.2023.08.003

23. Arian R, Hariri A, Mehridehnavi A, Fassihi A, Ghasemi F. Protein kinase inhibitors' classification using K-Nearest neighbor algorithm. *Comput Biol Chem*. 2020;86:107269. doi:10.1016/j.compbiolchem.2020.107269

24. Liao H, Zhang X, Zhao C, Chen Y, Zeng X, Li H. LightGBM: an efficient and accurate method for predicting pregnancy diseases. *J Obstet Gynaecol*. 2022;42(4):620–629. doi:10.1080/01443615.2021.1945006

25. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emergency Med*. 2011;18(10):1099–1104. doi:10.1111/j.1553-2712.2011.01185.x

26. Jinzaki M, Silverman SG, Akita H, Mikami S, Oya M. Diagnosis of renal angiomyolipomas: classic, fat-poor, and epithelioid types. *Semin Ultrasound CT MR*. 2017;38(1):37–46. doi:10.1053/j.sult.2016.11.001

27. Thiravit S, Teerasamit W, Thiravit P. The different faces of renal angiomyolipomas on radiologic imaging: a pictorial review. *British J Radiol*. 2018;91(1084). doi:10.1259/bjr.20170533

28. Lopes Vendrami C, Parada Villavicencio C, DeJulio TJ, et al. Differentiation of solid renal tumors with multiparametric MR imaging. *RadioGraphics*. 2017;37(7):2026–2042. doi:10.1148/rg.2017170039

29. Sun X-Y, Feng Q-X, Xu X, et al. Radiologic-radiomic machine learning models for differentiation of benign and malignant solid renal masses: comparison with expert-level radiologists. *Am J Roentgenol*. 2020;214(1):W44–W54. doi:10.2214/ajr.19.21617

30. Erdim C, Yardimci AH, Bektas CT, et al. Prediction of benign and malignant solid renal masses: machine learning-based CT texture analysis. *Acad Radiol*. 2020;27(10):1422–1429. doi:10.1016/j.acra.2019.12.015

31. Miskin N, Qin L, Silverman SG, Shinagare AB. Differentiating benign from malignant cystic renal masses: a feasibility study of computed tomography texture-based machine learning algorithms. *J Computer Assisted Tomograph*. 2023;47(3):376–381. doi:10.1097/rct.0000000000001433

32. Garnier C, Ferrer L, Vargas J, et al. A CT-based clinical, radiological and radiomic machine learning model for predicting malignancy of solid renal tumors (UroCCR-75). *Diagnostics*. 2023;13(15):2548. doi:10.3390/diagnostics13152548

33. Huang Y, Huang X, Wang A, et al. Individualized treatment decision model for inoperable elderly esophageal squamous cell carcinoma based on multi-modal data fusion. *BMC Med Inf Decis Making*. 2023;23(1). doi:10.1186/s12911-023-02339-5

34. Yusufaly TI, Zou J, Nelson TJ, et al. Improved prognosis of treatment failure in cervical cancer with nontumor PET/CT radiomics. *J Nucl Med*. 2022;63(7):1087–1093. doi:10.2967/jnumed.121.262618

35. Nassiri N, Maas M, Cacciamani G, et al. A radiomic-based machine learning algorithm to reliably differentiate benign renal masses from renal cell carcinoma. *European Urol Focus*. 2022;8(4):988–994. doi:10.1016/j.euf.2021.09.004

36. Ursprung S, Beer L, Bruining A, et al. Radiomics of computed tomography and magnetic resonance imaging in renal cell carcinoma-a systematic review and meta-analysis. *Eur Radiol*. 2020;30(6):3558–3566. doi:10.1007/s00330-020-06666-3

37. Posada Calderon LP, Eismann L, Reese SW, Reznik E, Hakimi AA. Advances in imaging-based biomarkers in renal cell carcinoma: a critical analysis of the current literature. *Cancers*. 2023;15(2):354. doi:10.3390/cancers15020354

38. Zheng Y, Zhou D, Liu H, Wen M. CT-based radiomics analysis of different machine learning models for differentiating benign and malignant parotid tumors. *Eur Radiol*. 2022;32(10):6953–6964. doi:10.1007/s00330-022-08830-3

39. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in alzheimer's disease: a systematic review. *Front Aging Neurosci*. 2017;9. 10.3389/fnagi.2017.00329