



Deep Learning for Classification of Solid Renal Parenchymal Tumors Using Contrast-Enhanced Ultrasound

Yun Bai¹ · Zi-Chen An¹ · Lian-Fang Du¹ · Fan Li² · Ying-Yu Cai¹

Received: 28 October 2024 / Revised: 22 April 2025 / Accepted: 24 April 2025 / Published online: 6 May 2025
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2025

Abstract

The purpose of this study is to assess the ability of deep learning models to classify different subtypes of solid renal parenchymal tumors using contrast-enhanced ultrasound (CEUS) images and to compare their classification performance. A retrospective study was conducted using CEUS images of 237 kidney tumors, including 46 angiomyolipomas (AML), 118 clear cell renal cell carcinomas (ccRCC), 48 papillary RCCs (pRCC), and 25 chromophobe RCCs (chRCC), collected from January 2017 to December 2019. Two deep learning models, based on the ResNet-18 and RepVGG architectures, were trained and validated to distinguish between these subtypes. The models' performance was assessed using sensitivity, specificity, positive predictive value, negative predictive value, F1 score, Matthews correlation coefficient, accuracy, area under the receiver operating characteristic curve (AUC), and confusion matrix analysis. Class activation mapping (CAM) was applied to visualize the specific regions that contributed to the models' predictions. The ResNet-18 and RepVGG-A0 models achieved an overall accuracy of 76.7% and 84.5% across all four subtypes. The AUCs for AML, ccRCC, pRCC, and chRCC were 0.832, 0.829, 0.806, and 0.795 for the ResNet-18 model, compared to 0.906, 0.911, 0.840, and 0.827 for the RepVGG-A0 model, respectively. The deep learning models could reliably differentiate between various histological subtypes of renal tumors using CEUS images in an objective and non-invasive manner.

Keywords Renal tumor · Contrast-enhanced ultrasound · Deep learning · Subtype classification

Introduction

The global incidence of kidney cancer has been on the rise. In 2020, kidney cancer ranked as the 9th most common cancer in men and the 14th most common in women [1]. Renal cell carcinoma (RCC) accounts for 90% of kidney cancers, with the majority being clear cell RCC (ccRCC, 70%), followed by papillary RCC (pRCC,

10–15%) and chromophobe RCC (chRCC, 5%) [2]. The histological subtype of RCC is an independent predictor of survival outcomes. Studies indicate that patients with ccRCC have a worse prognosis compared to those with pRCC and chRCC [3, 4]. Currently, ultrasound is the first-line imaging modality for screening and surveillance, while multiphasic contrast-enhanced computed tomography (CT) and magnetic resonance imaging (MRI) play pivotal roles in assessing renal lesions. Compared to CT or MRI, ultrasound offers several key features, such as high spatial and temporal resolution, enabling real-time examinations. Additionally, by using microbubble-structured agents, contrast-enhanced ultrasound (CEUS) can provide detailed quantification of tissue perfusion at the capillary level throughout all enhancement phases, without causing nephrotoxicity or exposing patients to radiation [5]. However, the imaging characteristics of different subtypes of renal tumors may overlap. For example, renal angiomyolipoma (AML) without visible fat exhibits early hyper-enhancement with contrast washout on contrast-enhanced CT and CEUS, mimicking the enhancement pattern of

Yun Bai and Zi-Chen An contributed equally to this article.

✉ Fan Li
medicinel@sjtu.edu.cn

✉ Ying-Yu Cai
yingyu.cai@shgh.cn

¹ Department of Ultrasound, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

² Department of Ultrasound, Shanghai Chest Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

ccRCC [6, 7]. ChRCC presents with slight hyperenhancement similar to that of the renal cortex, which can easily be mistaken for atypical ccRCC [7]. Therefore, it is challenging to determine the histological subtype of renal tumors using visual assessment alone, leading to a demand for more accurate, objective, and consistent classification methods.

Deep learning, a subset of machine learning, has revolutionized the field of diagnostic imaging by leveraging advanced algorithms to analyze and interpret complex medical images. As representative algorithms, convolutional neural networks (CNNs) are particularly effective for image classification, allowing for the identification of tumors [8]. Recent studies have demonstrated that CNNs perform comparably to or even better than radiologists in distinguishing between benign and malignant renal masses using CT or MRI images [9–12]. However, research on the diagnostic performance of CEUS is limited, and existing studies mainly focus on differentiating malignant from benign tumors rather than classifying RCC subtypes [13]. To our knowledge, there are no reports on the application of deep learning for the classification of subtypes of renal tumors based on CEUS images.

Therefore, the purpose of this study was to establish and evaluate deep learning models for the classification of solid renal parenchymal tumor subtypes based on CEUS images. By leveraging CNN-based approaches, we seek to enhance diagnostic accuracy and provide a more objective method for tumor subtype identification.

Materials and Methods

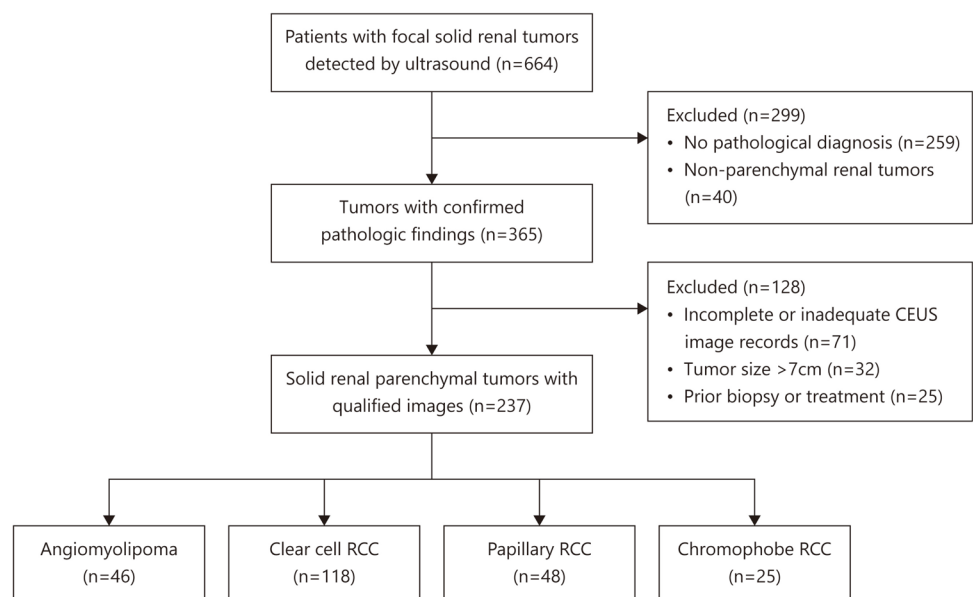
Patients

This retrospective study received approval from the local institutional ethics review board, which granted a waiver for patient consent due to the rigorous protection of anonymity. A review of our hospital's medical image database was performed from January 2017 to December 2019, and finally, a total of 237 patients were included in the study population. The inclusion criteria were (i) the presence of a focal solid renal tumor with a diameter of ≤ 7 cm on CEUS images; (ii) renal parenchymal neoplasms confirmed pathologically through surgery; (iii) complete CEUS images with adequate quality obtained within 1 month before surgery; and (iv) no prior localized or systemic treatment. In cases of bilateral tumors or multiple tumors in one kidney, the largest tumor was included. Tumors were categorized according to the 2022 World Health Organization classification of kidney tumors [14]. The flowchart for patient enrollment is shown in Fig. 1.

Contrast-Enhanced Ultrasound Imaging

CEUS examinations were conducted using one of two commercially available ultrasound systems: the Acuson Sequoia 512 (Siemens, Munich, Germany) with a 4 C1-S convex array transducer (frequency range, 1–4 MHz), or the LOGIQ E9 (GE Healthcare, Chicago, IL, USA) with

Fig. 1 Flowchart illustrating the enrollment process of renal tumor cases. CEUS, contrast-enhanced ultrasound; RCC, renal cell carcinoma



a C1-6 convex array transducer (frequency range, 1–6 MHz). After achieving optimal tumor visualization on grayscale ultrasound, the CEUS mode was activated. A contrast agent (SonoVue, Bracco, Milan, Italy) was administered intravenously in a dosage of 1.0–2.4 mL, adjusted according to the patient's weight, followed by a 5-mL flush of 0.9% saline solution. During CEUS imaging, the mechanical index was maintained at ≤ 0.23 to minimize exposure. Real-time video recording commenced at the start of the injection and was saved for at least 2 min in DICOM format.

Image Preprocessing

CEUS videos were converted into a series of images using RadiAnt DICOM Viewer software (version 2020.2.3; Medixant, Poznan, Poland). From each video, approximately 40–70 images capturing the largest tumor diameters during both the cortical and parenchymal phases were selected for subsequent normalization. A radiologist with 12 years of experience in abdominal CEUS manually annotated and cropped each image into a 12×12 cm square box, encompassing the tumor and surrounding renal parenchyma. The cropped image resolution was 353×353 pixels with a 24-bit depth. The radiologist was blinded to the histopathological data.

Deep Learning Model Development and Training

The final dataset was split into training, validation, and test sets in a ratio of 8:1:1. The deep learning model was developed and deployed using Python (version 3.7.12) with PyTorch (version 1.12.1). Given the limited size of the training set, the ResNet-18 and RepVGG-A0 networks were pre-trained on the ImageNet dataset before training, after which transfer learning was applied. To prevent overfitting, data augmentation techniques were employed, including horizontal flipping, contrast transformation, random rotation (± 1 degree), random resize ($\times 0.9$ to $\times 1.1$), and cropping to $224 \times 224 \times 3$ to fine-tune the pre-trained networks. The final models were trained using the stochastic gradient descent (SGD) optimizer on a computer equipped with a 12th Gen Intel Core i5-12500H CPU, an NVIDIA GeForce RTX 2050 GPU, and 16 GB of RAM. The learning rate, batch size, and number of epochs were set to 0.001, 16, and 100, respectively. Early stopping was applied when the validation loss did not decrease for ten consecutive epochs. To enhance the interpretability of the deep learning models, class activation mapping (CAM) was utilized to visualize the key areas contributing to the model's predictions [15]. The overview of the workflow of the study is shown in Fig. 2.

Statistical Analysis

Categorical variables among different subtypes of renal tumors were compared using a χ^2 test. Continuous variables were analyzed with a one-way analysis of variance (ANOVA), following the assessment of normal distributions by a one-sample Kolmogorov–Smirnov test. Diagnostic performance was assessed through metrics including sensitivity (recall), specificity, positive predictive value (precision), negative predictive value, F1 score, Matthews correlation coefficient (MCC), accuracy, area under the receiver operating characteristic (ROC) curve, and confusion matrix analysis. In the classification of a specific subtype, the other subtypes were treated as the negative class. For instance, in the ccRCC classification, AML, pRCC, and chRCC were considered as non-clear cell subtypes, and similarly for the other classifications.

Statistical analysis was conducted with SPSS statistics software (version 21) for Windows (IBM SPSS, Chicago, IL). A p -value < 0.05 was considered statistically significant. Average data are presented as mean $\pm$ SD.

Results

Patient and Image Characteristics

The demographic characteristics of the patients are presented in Table 1, while the distribution of renal tumor subtypes across different image sets is detailed in Table 2. A total of 237 patients (159 men and 78 women; mean age, 56.5 ± 8.6 years) with 237 renal lesions (mean size, 46.1 ± 9.9 mm) were included in the study. The cohort comprised 46 cases of AML, 118 cases of ccRCC, 48 cases of pRCC, and 25 cases of chRCC. There were no statistically significant differences in the distribution of gender, age, or tumor size among the subtypes. In terms of subtype composition, ccRCC accounted for the majority (49.8%, 118/237).

Performance of Deep Learning Models

Figure 3 presents the confusion matrices for the ResNet-18 and RepVGG-A0 models, illustrating the classification results for the four histological subtypes based on the test set. The overall accuracy across four subtypes in the test set was 76.7% for the ResNet-18 model and 84.5% for the RepVGG-A0 model, respectively. The results showed that both the ResNet-18 and RepVGG-A0 models achieved sensitivity greater than 80% for ccRCC. Notably, the RepVGG-A0 achieved over 80% sensitivity across all four subtypes. However, in comparison to the other three subtypes, both models performed suboptimally in recognizing chRCC. Both

Fig. 2 Overview of the workflow. CEUS, contrast-enhanced ultrasound; Conv., convolution; PPV, positive predictive value; NPV, negative predictive value; MCC, Matthews correlation coefficient; AUC, area under the curve; ROC, receiver operating characteristic; CAM, class activation map

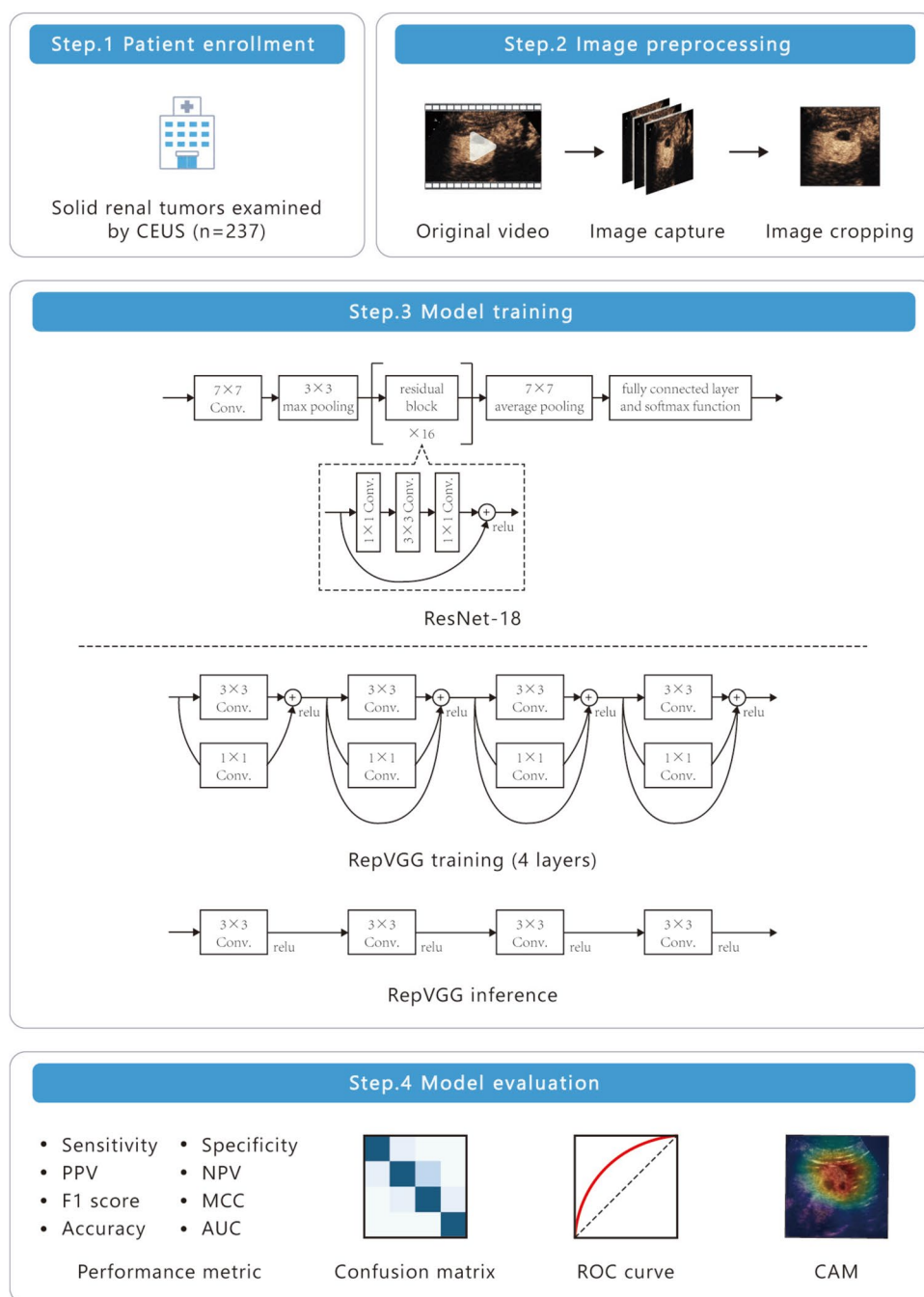


Table 1 Patient demographics

	Total	AML	ccRCC	pRCC	chRCC	<i>p</i>
Cases	237	46	118	48	25	
Gender						0.286
Male	67.1% (159/237)	16.4% (26/159)	53.5% (85/159)	19.5% (31/159)	10.7% (17/159)	
Female	32.9% (78/237)	25.6% (20/78)	42.3% (33/78)	21.8% (17/78)	10.3% (8/78)	
Age, range (years)	56.5 ± 8.6 (26–81)	54.2 ± 9.3 (37–77)	57.8 ± 8.6 (26–81)	56.5 ± 8.3 (41–72)	54.5 ± 6.9 (43–66)	0.056
Lesion size, range (mm)	46.1 ± 9.9 (22–70)	47.2 ± 10.3 (33–64)	45.6 ± 10.4 (22–70)	46.2 ± 8.9 (28–61)	46.1 ± 9.1 (32–63)	0.827

AML angiomyolipoma, ccRCC clear cell renal cell carcinoma, pRCC papillary renal cell carcinoma, chRCC chromophobe renal cell carcinoma

Table 2 Distribution of renal tumor subtypes across image sets

	Training set	Validation set	Test set	Total
AML	2099	244	252	2595
ccRCC	5308	679	650	6637
pRCC	2253	291	291	2835
chRCC	1506	182	202	1890

AML angiomyolipoma, ccRCC clear cell renal cell carcinoma, pRCC papillary renal cell carcinoma, chRCC chromophobe renal cell carcinoma

models exhibit good performance in identifying different subtypes of renal tumors, with the RepVGG-A0 model demonstrating superior results.

The performance metrics and area under the ROC curve (AUC) for the ResNet-18 and RepVGG-A0 models were calculated (Table 3). Both models achieved the AUC values over 0.8 for ccRCC, indicating that the CNN models exhibit strong diagnostic efficacy for this subtype. Notably, the AUC values of the RepVGG-A0 model exceeded 0.82 across all four histological subtypes of renal tumors, showcasing superior performance (Fig. 4).

Visual Evaluation of Deep Learning Models

CAM was applied to the trained CNNs to generate heat maps (Fig. 5). In the CAM maps of both CNN models, the tumor regions exhibited significant red activation, indicating that the models demonstrated high levels of interest in these regions. The focus areas identified by the ResNet-18 and

RepVGG-A0 models corresponded to the actual locations of the renal tumors, suggesting that deep learning can effectively distinguish between tumors and the surrounding renal parenchyma. Additionally, varying degrees of intensity and distribution of red activation were observed for different subtypes of renal tumors. Nevertheless, no obvious differences were noted among ccRCC, pRCC, and chRCC visually. This suggests that deep learning models may make predictions by detecting subtle morphological or internal features that are not readily observable to the naked eye.

Discussion

This study utilized the ResNet-18 and RepVGG-A0 models to classify four subtypes of solid renal parenchymal tumors using CEUS images. The ResNet-18 model achieved an overall classification accuracy of 76.7% for renal tumors, with a sensitivity of 82.9% for ccRCC and 68.0% for chRCC, a superior result compared to the same model applied to MRI. Zheng et al. [16] found that a deep CNN based on ResNet-18 had an overall accuracy of 60.4% in classifying renal tumor subtypes using a T2-weighted fat saturation MRI sequence. The RepVGG-A0 model achieved a sensitivity of 83.9% for AML and 88.0% for ccRCC, with an overall accuracy of 84.5%, comparable to the performance of the GoogLeNet network (accuracy = 85%), which used contrast-enhanced CT images to distinguish ccRCC, pRCC, and chRCC [17].

In this study, data imbalance was a concern, with a higher number of ccRCC (118 cases) compared to other subtypes.

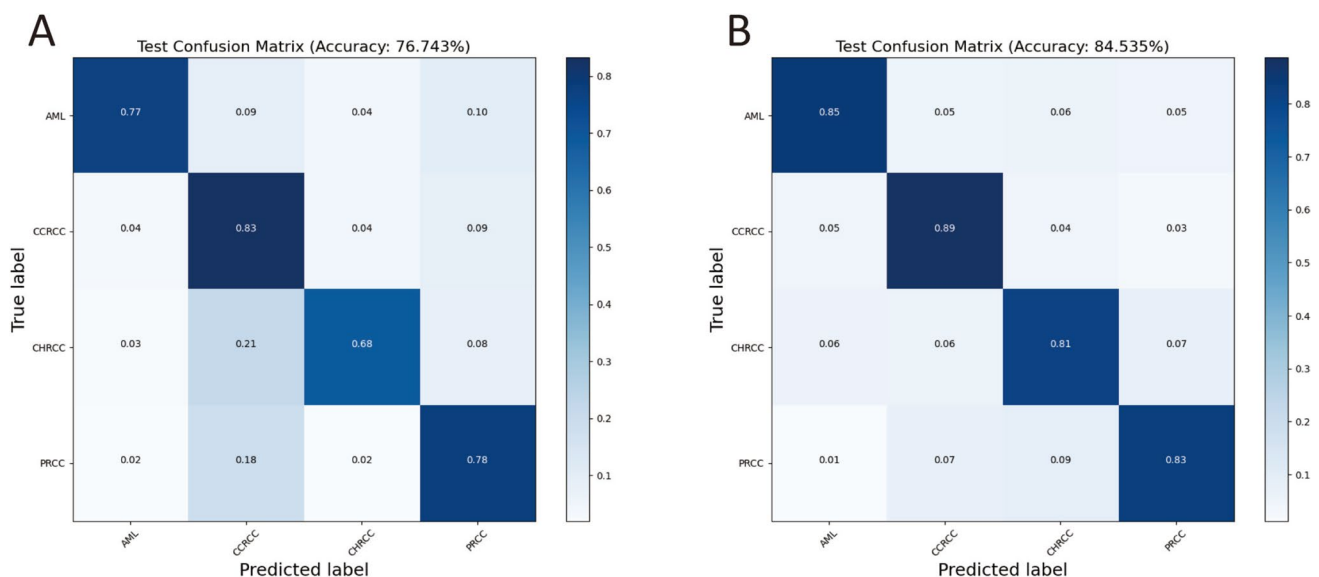


Fig. 3 Confusion matrices of the ResNet-18 model (A) and the RepVGG-A0 model (B) in classifying renal tumor subtypes. AML, angiomyolipoma; ccRCC, clear cell renal cell carcinoma; chRCC, chromophobe renal cell carcinoma; pRCC, papillary renal cell carcinoma

Table 3 Performance of deep learning models in identifying renal tumors

Models	Subtypes	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	F1 score (95% CI)	MCC (95% CI)	Accuracy (95% CI)	AUC (95% CI)
ResNet-18	AML	0.770 (0.718–0.821)	0.966 (0.955–0.976)	0.833 (0.785–0.880)	0.950 (0.937–0.962)	0.800 (0.754–0.846)	0.759 (0.736–0.781)	0.931 (0.917–0.944)	0.832 (0.814–0.850)
	ccRCC	0.829 (0.800–0.857)	0.839 (0.812–0.864)	0.818 (0.788–0.847)	0.849 (0.823–0.875)	0.824 (0.797–0.851)	0.668 (0.643–0.693)	0.835 (0.815–0.854)	0.829 (0.807–0.846)
	pRCC	0.782 (0.723–0.837)	0.910 (0.894–0.926)	0.596 (0.536–0.657)	0.961 (0.950–0.972)	0.677 (0.625–0.729)	0.621 (0.596–0.647)	0.892 (0.876–0.908)	0.806 (0.784–0.823)
	chRCC	0.680 (0.625–0.732)	0.964 (0.952–0.975)	0.832 (0.782–0.878)	0.920 (0.903–0.935)	0.749 (0.702–0.795)	0.696 (0.672–0.720)	0.905 (0.889–0.920)	0.795 (0.772–0.815)
RepVGG-A0	AML	0.839 (0.792–0.882)	0.955 (0.943–0.966)	0.805 (0.756–0.850)	0.964 (0.953–0.975)	0.821 (0.779–0.864)	0.781 (0.760–0.803)	0.934 (0.921–0.947)	0.906 (0.884–0.921)
	ccRCC	0.880 (0.854–0.904)	0.941 (0.924–0.957)	0.929 (0.909–0.949)	0.899 (0.877–0.919)	0.904 (0.882–0.926)	0.825 (0.805–0.845)	0.912 (0.898–0.927)	0.911 (0.895–0.928)
	pRCC	0.832 (0.777–0.881)	0.956 (0.944–0.967)	0.760 (0.701–0.814)	0.971 (0.961–0.981)	0.794 (0.745–0.844)	0.759 (0.737–0.781)	0.938 (0.925–0.951)	0.840 (0.828–0.857)
	chRCC	0.814 (0.769–0.859)	0.947 (0.934–0.960)	0.800 (0.753–0.844)	0.951 (0.939–0.964)	0.807 (0.765–0.848)	0.756 (0.734–0.779)	0.920 (0.905–0.934)	0.827 (0.803–0.853)

AML angiomyolipoma, ccRCC clear cell renal cell carcinoma, pRCC papillary renal cell carcinoma, chRCC chromophobe renal cell carcinoma, PPV positive predictive value, NPV negative predictive value, MCC Matthews correlation coefficient, AUC area under the curve, CI confidence interval

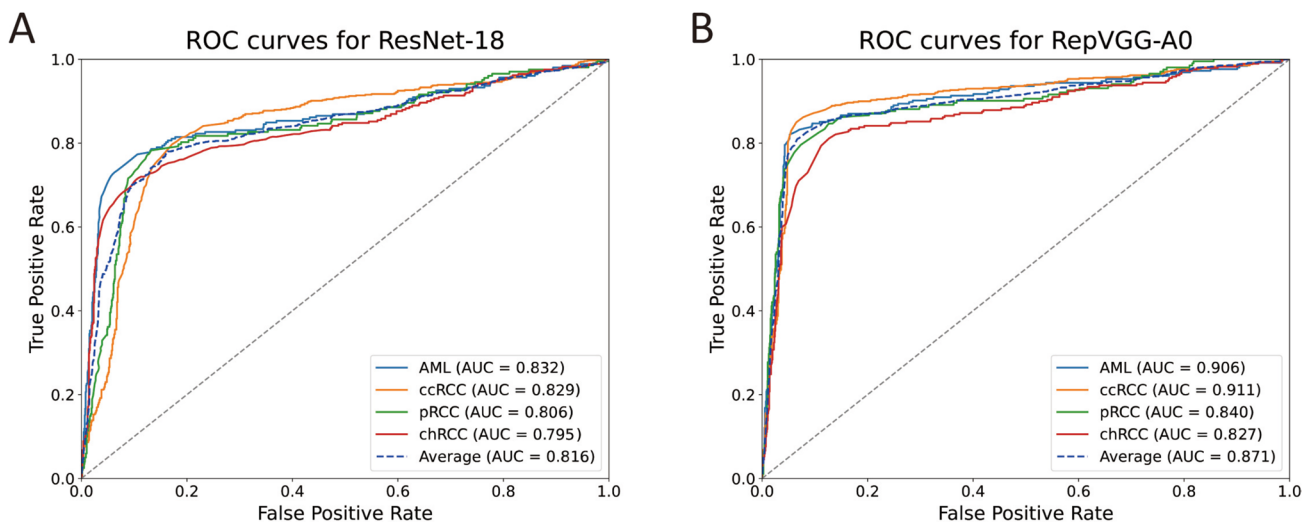


Fig. 4 The receiver operating characteristic (ROC) curves of the ResNet-18 model (A) and the RepVGG-A0 model (B). AML, angiomyolipoma; ccRCC, clear cell renal cell carcinoma; pRCC, papillary

renal cell carcinoma; chRCC, chromophobe renal cell carcinoma; AUC, area under the curve

To address this, we did not rely solely on accuracy to evaluate classification performance but also employed the F1 score and MCC, which provide more comprehensive

assessments, particularly for imbalanced datasets. The F1 score balances precision and recall, while the MCC considers all elements of the confusion matrix, making it a

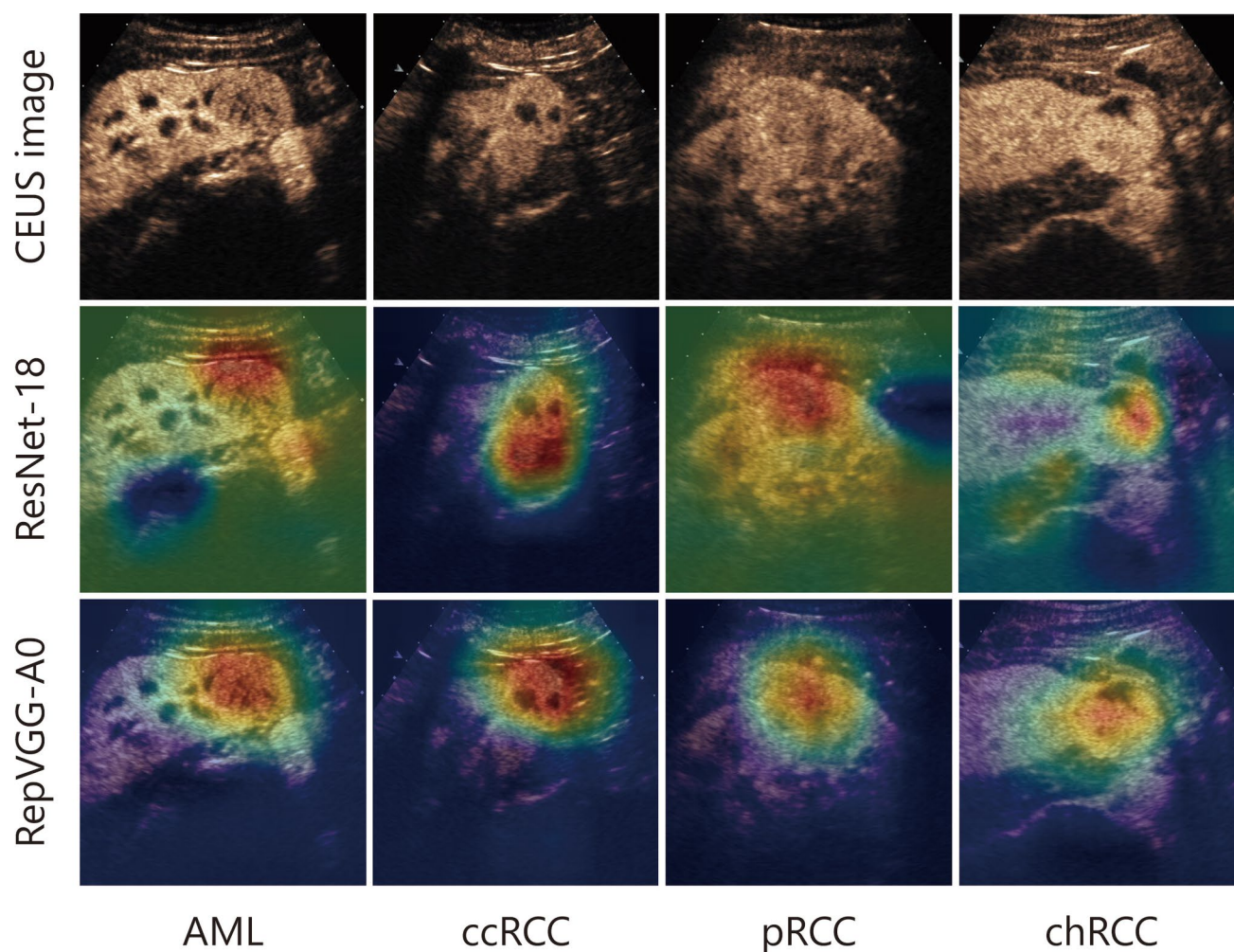


Fig. 5 Class activation maps (CAM) to visualize the areas where the deep learning models focused most. The red-highlighted regions of the tumors indicate that the models assigned higher weights to those specific areas during classification. CEUS, contrast-enhanced ultra-

sound; AML, angiomyolipoma; ccRCC, clear cell renal cell carcinoma; pRCC, papillary renal cell carcinoma; chRCC, chromophobe renal cell carcinoma

more reliable metric for evaluating classification performance across uneven class distributions [18, 19]. Using the RepVGG-A0 model for subtype classification of renal tumors, the accuracy for AML, ccRCC, pRCC, and chRCC was similar (93.4%, 91.2%, 93.8%, and 92.0%, respectively). However, differences were observed in the F1 score (0.821, 0.904, 0.794, 0.807) and MCC (0.781, 0.825, 0.759, 0.756). Based on these metrics, ccRCC exhibited the highest classification performance, while pRCC and chRCC showed lower effectiveness. This discrepancy may be attributed to two factors. First, due to their histological characteristics, pRCC and chRCC tend to exhibit homogeneous hypoenhancement on CEUS, making them more challenging to distinguish from each other or from AML [20–23], which may limit the model's ability to differentiate these tumor types. Second, the relatively smaller sample sizes of pRCC (48 cases) and chRCC (25 cases) compared to ccRCC may have also limited

the model's ability to learn distinctive features, contributing to the reduced classification performance. Expanding the dataset in future studies is expected to improve the models' capability.

Our study demonstrates that the RepVGG-A0 model outperformed the ResNet-18 model, achieving higher AUCs across all subtypes of renal tumors (AML, 0.906 vs 0.832; ccRCC, 0.911 vs 0.829; pRCC, 0.840 vs 0.806; chRCC, 0.827 vs 0.795). ResNet is a multi-branch network composed of residual blocks designed to mitigate the degradation problem in deep networks, allowing for more efficient learning of complex features [24]. However, its multi-branch structure increases memory consumption, as feature maps from each branch occupy GPU memory during both training and inference, which can limit efficiency, particularly in resource-constrained environments [25]. In contrast, RepVGG adopts a multi-branch topology during

training but transforms into a VGG-like plain architecture during inference. This design not only reduces memory overhead but also accelerates inference, making it more computationally efficient. Additionally, the structural reparameterization in RepVGG enhances learning efficiency and reduces the risk of overfitting, particularly in scenarios with limited training data [25]. Our study demonstrated that when classifying CEUS images in smaller datasets, a less complex and shallower architecture like RepVGG may offer a better balance between performance and efficiency compared to deeper residual networks. These findings highlight the importance of selecting an appropriate model architecture based on dataset size and clinical application constraints.

Ultrasound is an essential examination for detecting suspicious renal lesions and differentiating solid from cystic renal masses [26]. Furthermore, CEUS offers real-time perfusion characteristics of renal tumors in both the arterial and venous phases. This makes it a valuable supplemental tool for distinguishing between benign and malignant renal tumors, especially in patients with allergies to CT contrast agents or those with renal insufficiency. Although some studies have applied deep learning models to distinguish benign from malignant renal lesions using CT or MRI, achieving accuracy rates comparable to radiologists (77.3 ~ 86.5%), research on the application of CNNs to CEUS images remains limited [9–11, 27]. Zhu et al. [13] developed a multimodal network using EfficientNet as the backbone, incorporating grayscale ultrasound and CEUS to classify benign and malignant solid renal tumors, achieving an accuracy of 80.0% and an AUC of 0.877; this performance surpassed that of a senior radiologist. Notably, previous studies have shown that it remains challenging to differentiate between pRCC and chRCC based on CEUS characteristics assessed by radiologists or even with the assistance of quantitative analysis software [28, 29]. In contrast, our study demonstrates that deep learning models can effectively distinguish between these subtypes using CEUS images, highlighting the potential of artificial intelligence (AI) to capture subtle imaging features beyond human perception. This underscores the added value of AI-assisted diagnostic tools in improving the accuracy of renal tumor subtype classification, particularly for subtypes that are visually similar on CEUS. Such capabilities could contribute to more precise preoperative planning and personalized treatment strategies. Building on these findings, we believe that the integration of deep learning networks could further advance the diagnostic utility of CEUS in renal tumors, especially as a viable alternative for patients with contraindications to contrast-enhanced CT or MRI.

There are several limitations in this study. First, only CEUS features were extracted, and the value of grayscale

ultrasound images was not verified. Since B-mode ultrasound provides valuable information on tissue echogenicity and echotexture through its gray shades, incorporating it may enhance the classification performance. Second, the number of ccRCC cases included in this study exceeded that of other subtypes, which may have resulted in an imbalanced dataset and potentially led to poorer performance on the minority subtypes. Third, this is a single-center study and lacks an external independent test set. Further multicenter validation and expansion of larger datasets are needed before it can be applied in clinical practice.

Conclusion

In this study, we employed deep learning models based on ResNet-18 and RepVGG-A0 architectures to classify CEUS images of AML, ccRCC, pRCC, and chRCC with satisfactory accuracy. Deep learning models can effectively differentiate renal tumor subtypes using CEUS images, achieving accuracy comparable to or exceeding that of MRI- and CT-based studies. The RepVGG-A0 model outperformed ResNet-18 in both discriminability and robustness, demonstrating the potential advantages of structural re-parameterization in deep learning for medical imaging.

Future studies should focus on validating these findings with larger, multicenter datasets, optimizing model architectures for improved generalizability, and exploring the integration of deep learning with radiomics or multimodal imaging to enhance diagnostic performance. These efforts could help translate deep learning-based CEUS analysis into clinical practice, ultimately improving the accuracy and efficiency of renal tumor diagnosis.

Author Contribution Y Bai: manuscript writing, data collection, data analysis. ZC An: manuscript writing, data analysis, technical support. LF Du: data analysis, data collection. F Li: project development, manuscript editing. YY Cai: project development, data analysis, manuscript editing.

Funding The authors gratefully acknowledge the support of Shanghai Jiao Tong University K. C. Wong Medical Fellowship Fund and Shanghai Municipal Education Commission.

Data Availability The deidentified data supporting the findings of this study is available from the corresponding authors upon reasonable request. However, due to privacy and ethical considerations, the data is not publicly accessible.

Declarations

Ethics Approval The local institutional review board approved this study.

Competing Interests The authors declare no competing interests.

References

1. Bukavina L, Bensalah K, Bray F, et al. Epidemiology of Renal Cell Carcinoma: 2022 Update. *Eur Urol* 82: 529–542, 2022.
2. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. *Eur Urol* 70: 93–105, 2016.
3. Patard JJ, Leray E, Rioux-Leclercq N, et al. Prognostic value of histologic subtypes in renal cell carcinoma: a multicenter experience. *J Clin Oncol* 23: 2763–2771, 2005.
4. Leibovich BC, Lohse CM, Crispen PL, et al. Histological subtype is an independent predictor of outcome for patients with renal cell carcinoma. *J Urol* 183: 1309–1315, 2010.
5. Dietrich CF, Correas JM, Cui XW, et al. EFSUMB Technical Review - Update 2023: Dynamic Contrast-Enhanced Ultrasound (DCE-CEUS) for the Quantification of Tumor Perfusion. *Ultraschall Med* 45: 36–46, 2024.
6. Lim RS, Flood TA, McInnes MDF, Lavallee LT, Schieda N. Renal angiomyolipoma without visible fat: Can we make the diagnosis using CT and MRI? *Eur Radiol* 28: 542–553, 2018.
7. King KG. Use of Contrast Ultrasound for Renal Mass Evaluation. *Radiol Clin North Am* 58: 935–949, 2020.
8. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer* 18: 500–510, 2018.
9. Dai C, Xiong Y, Zhu P, et al. Deep Learning Assessment of Small Renal Masses at Contrast-enhanced Multiphase CT. *Radiology* 311: e232178, 2024.
10. Xu Q, Zhu Q, Liu H, et al. Differentiating Benign from Malignant Renal Tumors Using T2- and Diffusion-Weighted Images: A Comparison of Deep Learning and Radiomics Models Versus Assessment from Radiologists. *J Magn Reson Imaging* 55: 1251–1259, 2022.
11. Yao H, Tian L, Liu X, et al. Development and external validation of the multichannel deep learning model based on unenhanced CT for differentiating fat-poor angiomyolipoma from renal cell carcinoma: a two-center retrospective study. *J Cancer Res Clin Oncol* 149: 15827–15838, 2023.
12. Zuo T, Zheng Y, He L, et al. Automated Classification of Papillary Renal Cell Carcinoma and Chromophobe Renal Cell Carcinoma Based on a Small Computed Tomography Imaging Dataset Using Deep Learning. *Front Oncol* 11: 746750, 2021.
13. Zhu D, Li J, Li Y, et al. Multimodal ultrasound fusion network for differentiating between benign and malignant solid renal tumors. *Front Mol Biosci* 9: 982703, 2022.
14. Moch H, Amin MB, Berney DM, et al. The 2022 World Health Organization Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. *Eur Urol* 82: 458–468, 2022.
15. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*: 2921–2929, 2016.
16. Zheng Y, Wang S, Chen Y, Du HQ. Deep learning with a convolutional neural network model to differentiate renal parenchymal tumors: a preliminary study. *Abdom Radiol (NY)* 46: 3260–3268, 2021.
17. Hwang SI, Han S, Lee HJ. The Classification of Renal Cancer in 3-Phase CT Images Using a Deep Learning Method. *J Digit Imaging* 32: 638–643, 2019.
18. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21: 6, 2020.
19. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 45: 427–437, 2009.
20. Cai Y, Li F, Li Z, et al. Predictive value of contrast-enhanced ultrasound combined with conventional ultrasound in solid renal parenchymal lesions. *Br J Radiol* 94: 20210518, 2021.
21. Siracusano S, Bertolotto M, Ciciliato S, Valentino M, Liguori G, Visalli F. The current role of contrast-enhanced ultrasound (CEUS) imaging in the evaluation of renal pathology. *World J Urol* 29: 633–638, 2011.
22. Sun D, Wei C, Li Y, Lu Q, Zhang W, Hu B. Contrast-Enhanced Ultrasonography with Quantitative Analysis allows Differentiation of Renal Tumor Histotypes. *Sci Rep* 6: 35081, 2016.
23. Zhang Q, Wei S, Yang B, Qian X. Differences Between Type I and Type II Papillary Renal Cell Carcinoma on Ultrasound. *Advanced Ultrasound in Diagnosis and Therapy* 3: 187–192, 2019.
24. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 2016: 770–778, 2016.
25. Ding X, Zhang X, Ma N, Han J, Ding G, Sun J. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*: 13733–13742, 2021.
26. Escudier B, Porta C, Schmidinger M, et al. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 30: 706–720, 2019.
27. Nikpanah M, Xu Z, Jin D, et al. A deep-learning based artificial intelligence (AI) approach for differentiation of clear cell renal cell carcinoma from oncocytoma on multi-phasic MRI. *Clin Imaging* 77: 291–298, 2021.
28. Cao H, Fang L, Chen L, et al. The Value of Contrast-Enhanced Ultrasound in Diagnosing Small Renal Cell Carcinoma Subtypes and Angiomyolipoma. *J Ultrasound Med* 41: 1415–1423, 2022.
29. Li CX, Lu Q, Huang BJ, et al. Quantitative evaluation of contrast-enhanced ultrasound for differentiation of renal cell carcinoma subtypes and angiomyolipoma. *Eur J Radiol* 85: 795–802, 2016.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.