An End-to-End Deep Learning Framework and Multicenter Validation for the Diagnosis of Retroperitoneal Soft Tissue Sarcoma

Tong Meng[1,2†], Rui Chen[3], Xiang Feng[4] , Qianwen Zhang[4] , Yue Yang[5] , Jianbo Gao[6], Lian Yang[7]

[1] Department of Orthopedics, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China.

[2] Tongji University Cancer Center, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai, China.

[3] Department of Urology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

[4] Department of Urology, Shanghai Changhai Hospital, Naval Medical University, Shanghai, China.

[5] Department of Nephrology, First Medical Center of Chinese PLA General Hospital, Beijing, China.

[6] Department of Radiology, The First Affiliated Hospital of Zhengzhou University, No.1 Jianshe Road, Zhengzhou, Henan, China.

[7] Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.

Background. Retroperitoneal soft tissue sarcomas (RSTSs) comprise a heterogeneous group of tumors that remain diagnostically challenging even relying on histopathology. Preoperative diagnosis is a key factor in determining the therapeutic strategy and postoperative prognosis of patients. Computed tomography (CT) is a potential auxiliary diagnostic method, yet a multicenter-validated diagnostic tool is lacking for RSTSs. This study aimed to develop and validate an end-to-end deep learning model, RESIND (REtroperitoneal soft tissue Sarcomas artificial-INtelligence Diagnosis), for the accurate diagnosis and segmentation of RSTS using enhanced CT images.

Methods. This multicenter study involved patients recruited from 12 Chinese centers between January 2012 and June 2024. All participants had histologically confirmed RSTSs, encompassing seven subtypes: dedifferentiated liposarcoma, well-differentiated liposarcoma, leiomyosarcoma, ganglioneuroma, lymphoma, schwannoma, and paraganglioma. The RESIND model was initially developed using retrospective data from a single center (n = 606), with five-fold cross-validation for internal validation. Its performance was then externally tested on a retrospective cohort from 11 different hospitals (n = 736). Finally, a prospective validation was conducted on a cohort from the same primary hospital as the training set, enrolled between January and June 2024 (n = 188). To evaluate its clinical utility, we further performed a reader study involving 30 radiologists from 11 hospitals across China.

Results. RESIND demonstrated robust and consistent predictive performance across all cohorts. In the seven-way classification task, the model achieved top-1 accuracies of 0.66 (95% CI: 0.61–0.69) in the training cohort, 0.61 (95% CI: 0.46–0.73) in the external validation cohort, and 0.63 (95% CI: 0.54–0.69) in the prospective validation cohort. Corresponding top-2 accuracies were 0.82 (95% CI: 0.79–0.85), 0.79 (95% CI: 0.77–0.83), and 0.77 (95% CI: 0.71–0.83), respectively. For most sarcoma types, the area under the ROC curve (AUC) exceeded 0.80. In the segmentation task, RESIND yielded average Dice scores of 0.75 (95% CI: 0.73–0.76), 0.72 (95% CI: 0.70–0.74), and 0.73 (95% CI: 0.70–0.77) in the training, external validation, and prospective validation cohorts, respectively. In the reader study, RESIND's top-1 accuracy (64.0%) significantly surpassed that of junior (42.6%, p = 0.006) and attending

radiologists (57.4%, p = 0.009), and was comparable to that of senior radiologists (64.3%, p = 0.905). When assisted by RESIND, the diagnostic accuracy of both junior and attending radiologists improved markedly. Furthermore, RESIND assistance significantly reduced interpretation time and increased diagnostic certainty across all radiologist groups.

Conclusion. RESIND is a first-in-class, multi-center validated model for the simultaneous diagnosis and segmentation of RSTSs. By demonstrating the potential to improve diagnostic accuracy, increase diagnostic certainty, and reduce interpretation time, this study underscores the clinical promise of artificial intelligence in supporting radiologists managing these rare and complex tumors.