Respiratory Research

RESEARCH

Open Access

Prediction of tumor spread through air spaces with an automatic segmentation deep learning model in peripheral stage I lung adenocarcinoma

Cong Liu[1], Yu-feng Wang[2], Ping Gong[3*], Xiu-Qing Xue[4], Hong-Ying Zhao[5], Hui Qian[6*], Chao Jia[7] and Xiao-Feng Li[7*]

Abstract

Background To evaluate the clinical applicability of deep learning (DL) models based on automatic segmentation in preoperatively predicting tumor spread through air spaces (STAS) in peripheral stage I lung adenocarcinoma (LUAD).

Methods This retrospective study analyzed data from patients who underwent surgical treatment for lung tumors from January 2022 to December 2023. An external validation set was introduced to assess the model's generalizability. The study utilized conventional radiomic features and DL models for comparison. ROI segmentation was performed using the VNet architecture, and DL models were developed with transfer learning and optimization techniques. We assessed the diagnostic accuracy of our models via calibration curves, decision curve analysis, and ROC curves.

Results The DL model based on automatic segmentation achieved an AUC of 0.880 (95% CI 0.780–0.979), outperforming the conventional radiomics model with an AUC of 0.833 (95% CI 0.707–0.960). The DL model demonstrated superior performance in both internal validation and external testing cohorts. Calibration curves, decision curve analysis, and ROC curves confirmed the enhanced diagnostic accuracy and clinical utility of the DL approach.

Conclusion The DL model based on automatic segmentation technology shows significant promise in preoperatively predicting STAS in peripheral stage I LUAD, surpassing traditional radiomics models in diagnostic accuracy and clinical applicability.

Clinical trial number The clinical trial was registered on April 22, 2024, with the registration number researchregistry10213 (www.researchregistry.com).

Keywords Radiomics, Deep Learning, Lung Adenocarcinoma, Tumor Spread through Air Spaces

[†]Cong Liu and Xiu-Qing Xue both are co-first authors.

*Correspondence:
Ping Gong
gongping@xzhmu.edu.cn
Hui Qian
1stmmm1st@163.com
Xiao-Feng Li
lxf5818@163.com
Full list of author information is available at the end of the article

Introduction

In 2015, the World Health Organization introduced the concept of tumor spread through air spaces (STAS) in its lung cancer classification [1]. Subsequent studies have confirmed that STAS is an independent risk factor for recurrence in patients with stage I lung adenocarcinoma (LUAD) who undergo sublobar resection [2]. Eguchi [3] suggested that for patients with T1 stage LUAD who are STAS-positive, lobectomy offers greater survival benefits compared to sublobar resection. Furthermore, STAS is also an independent adverse prognostic factor for patients with stage I LUAD [4, 5], significantly associated with recurrence-free survival [6]. Therefore, accurate preoperative identification of STAS is critical for surgical planning and prognostic evaluation in stage I LUAD.

Current studies indicate that intraoperative frozen section (FS) analysis has a sensitivity of 50% and a negative predictive value of only 8%, rendering it suboptimal for diagnosing STAS [7]. The limited efficacy of intraoperative FS diagnosis of STAS can affect the extent of resection and the choice of surgical method [8, 9]. Additionally, due to the difficulty in obtaining live tissue specimens for pathological diagnosis of tumor cells within alveolar or air spaces, preoperative percutaneous biopsy is also inadequate for definitive STAS diagnosis. Thus, there is an urgent need for a more accurate preoperative method to diagnose STAS.

Recently, imaging-based deep learning (DL) tools in the computer vision field have gained significant attention. These tools have shown great promise in quantifying early-stage lung cancer heterogeneity and providing potential clinical imaging features for patient stratification. For example, radiomics-based clinical malignancy probability assessments have demonstrated considerable potential [10, 11]. Accurate tumor delineation is a priority in radiomics; however, challenges remain regarding the accuracy and reproducibility of early-stage lung cancer lesion delineation and the robustness of radiomic feature extraction [12].

To address these challenges, deep convolutional neural networks (CNNs) have achieved significant success in medical image segmentation. CT images, being volumetric data, require full utilization of spatial information [13, 14]. Additionally, two major challenges in the field include: (1) label scarcity due to the cost of annotations by experienced domain experts, and (2) the higher risk of overfitting due to increased parameter numbers.

This study proposes a novel deep learning framework combining automatic segmentation and prediction models to address these issues. Specifically, we employ the VNet architecture for automatic ROI segmentation and ResNet-based models with transfer learning for STAS prediction. Our approach evaluates the diagnostic performance of deep learning models in comparison to conventional radiomics-based models. We hypothesize that the deep learning models, when coupled with automatic segmentation techniques, will outperform radiomics models in terms of diagnostic accuracy and clinical applicability.

The primary aim of this study is to explore the feasibility and accuracy of using automatic segmentation combined with deep learning to predict STAS preoperatively in peripheral stage I LUAD. This study also evaluates the clinical utility of combining these techniques to provide a non-invasive, reproducible, and efficient diagnostic workflow for thoracic oncology applications.

Materials and methods

Study design and dataset

This retrospective study analyzed data collected from January 2022 to December 2023. Clinical and radiological data were obtained from patients who underwent surgical treatment for lung tumors at our institution, supplemented with an external validation set from another hospital. Inclusion criteria were as follows: (i) clinical stage T1-T2aN0M0 according to the 8th edition of the American Joint Committee on Cancer cancer staging manual [15]; (ii) tumors located in the outer two-thirds of the lung field on chest CT axial images, with the tumor center within this specified area; (iii) radical resection for lung cancer and systematic lymph node dissection with a minimum of six lymph nodes excised; (iv) postoperative pathological diagnosis confirmed as adenocarcinoma. Exclusion criteria included: (i) multiple pulmonary neoplastic lesions diagnosed preoperatively or synchronous primary or multiple primary lung cancers (more than two lesions) identified postoperatively; (ii) preoperative exposure to radiotherapy, chemotherapy, immunotherapy, or targeted therapy for cancer; (iii) a history of other malignant tumors within the past three years.

The study received approval from the local Institutional Review Board (2023–02–027-K01) and adhered to the Declaration of Helsinki. Informed consents were waived by the Committee due to the retrospective and anonymous nature of this study. The study was registered in the Research Registry (researchregistry10213). Compliance with the CheckList for EvaluAtion of Radiomics research (CLEAR) guidelines was maintained [16, 17].

Image preprocessing and segmentation

All CT scans were performed using GE Discovery 750HD, SIEMENS SOMATOM Definition AS, and SOMATOM Definition Flash scanners, spanning from the apex to the base of the lungs. Patients were positioned supine, with scan parameters set at a tube voltage of 120 kV and an automatic tube current ranging from

80 to 350 mA. The standard scanning slice thickness and interval were 5 mm, with a reconstructed slice thickness and interval of 0.6–0.625 mm. Images were analyzed using both lung (window width 1500 HU, window level − 450 HU) and mediastinal (window width 350 HU, window level 35 HU) settings.

For segmentation, the VNet architecture was employed to automatically delineate regions of interest (ROIs) within CT volumes. The VNet model was trained for 300 epochs using a batch size of 16, the Adam optimizer with a learning rate of 0.001, and the Dice loss function. Each layer consisted of 3D convolutional operations with kernel sizes of $3\times3\times3$, followed by ReLU activation and batch normalization. An early stopping mechanism was implemented to preserve the most efficient model configurations. Predicted ROIs were validated using metrics such as the Dice similarity coefficient (DSC) and Intersection over Union (IoU). Visualization of the segmentation results, including ground truth markings, is shown in the results section with corresponding metrics. The detailed methodologies utilized for training are described in Supplementary Material 1A.

Radiomics feature extraction and model construction

Handcrafted radiomic features were extracted using Pyradiomics (http://pyradiomics.readthedocs.io) [18]. These features included geometric, intensity, and texture-based attributes, such as gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size zone matrix (GLSZM), and neighborhood gray-tone difference matrix (NGTDM). A total of 1834 features were extracted and z-score normalized. Features were filtered based on statistical significance ($p<0.05$) and Pearson's correlation coefficient (<0.9). LASSO regression was then applied to construct the radiomics signature, selecting features through tenfold cross-validation. The model achieving the highest performance on the validation cohort was chosen for comparison. For the radiomics approach, features were extracted from all slices containing the tumor, and the features were aggregated to obtain a case-wise score.

Deep learning framework

The deep learning framework was designed using the ResNet architecture with transfer learning. ROI volumes generated by the VNet segmentation were first resampled to a uniform spatial resolution of $0.625\times0.625\times0.625$ mm^3 using trilinear interpolation. The interpolated volumes were then cropped to a fixed size of [96, 96, 96] voxels centered on the tumor centroid, ensuring preservation of spatial relationships. Intensity normalization was performed by z-scoring ($\mu=0$, $\sigma=1$) across the entire volume. To improve robustness against segmentation variability, random cropping was implemented with dynamic region calculation ($\text{Crop\_size}=\text{ROI}_{\text{diameter}}+2\times\text{Safety\_margin}$) where the safety margin was set to 10 pixels. When the calculated crop region exceeded image boundaries, mirror padding was applied to avoid information loss. The tumor centroid was guaranteed to remain within the cropped region through coordinate constraints.

For the deep learning approach, only the slice with the largest ROI was used for prediction. Data augmentation included random rotations, flips, and scaling. Fusion approaches (min, max, mean) were implemented and assessed for their impact on predictive performance, as detailed in the results section.

Hyperparameter optimization focused on the learning rate, batch size (32), and number of epochs. A cosine decay learning rate schedule was applied to stabilize training and improve generalization (Supplementary Table S1). Models were trained using Stochastic Gradient Descent (SGD) with softmax cross-entropy loss.

$$\eta_t = \eta_{min}^i + \frac{1}{2}\left(\eta_{max}^i - \eta_{min}^i\right)\left(1+\cos\left(\frac{T_{cur}}{T_i}\pi\right)\right)$$

Here, $\eta_{\text{min}}^{\text{i}} = 0$ is the minimum learning rate, $\eta_{\text{max}}^{\text{i}} = 0.01$ is the maximum learning rate, and $\text{T}_{\text{i}} = 30$ is the cycle length for each epoch. We also selected Stochastic Gradient Descent (SGD) for optimization and employed softmax cross-entropy for loss calculation.

Model evaluation metrics

Model performance was evaluated using receiver operating characteristic (ROC) curves, area under the curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Calibration curves and the Hosmer–Lemeshow test were used to assess model calibration. Decision curve analysis (DCA) evaluated the clinical utility of each model.

Statistical methodology

Our statistical evaluations and model development were executed using Python version 3.7.12, supplemented by the statsmodels library version 0.13.2. Machine learning frameworks were developed employing the scikit-learn library version 1.0.2. DL training utilized an NVIDIA 4090 GPU, with software frameworks including MONAI version 0.8.1 and PyTorch version 1.8.1. The segmentation of 3D regions in the training set was conducted using the 3D Slicer software (version 5.3.0–2023-08–03).

Results

Baseline characteristics

A total of 290 cases met the inclusion and exclusion criteria, with 65 cases (22.41%) testing positive for STAS.

The cohort comprised 55% males and 45% females, with an average age of 62 years (SD: ± 9.3 years). No statistically significant differences were found between clinical and pathological variables from Center 1 and Center 2, as all p-values exceeded 0.05. These findings suggest that the two centers were comparable and suitable for pooled analysis. The study enrollment process is depicted in Figs. 1, 2 outlines the detailed study flowchart.

Radiomics signature

Feature statistics

A total of 1834 handcrafted features were extracted, categorized into six groups: first-order statistics (360 features), shape-based features (14 features), and texture features, which included gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size zone matrix (GLSZM), and neighborhood gray-tone difference matrix (NGTDM). All features were extracted using an in-house feature analysis program built on Pyradiomics (http://pyradiomics.readthedocs.io) [18]. Figure 3 presents an overview of the extracted features along with their p-value distribution, highlighting features significantly associated with STAS.

Lasso feature selection

To construct the radiomics signature, Lasso regression was applied, resulting in the selection of 12 nonzero coefficients that were used to calculate the Rad-score. The process incorporated tenfold cross-validation to ensure model robustness, and the mean standard error (MSE) curve is shown in Fig. 4. The radiomics-based XGBoost model achieved an Area Under the Curve (AUC) of 0.833 (95% CI 0.707–0.960) in the validation cohort. While these results indicate moderate discriminative capability, the AUC was slightly lower compared to DL-based methods. Figures 5 and 6 provide detailed performance metrics and visualizations for this model.

Deep learning signature

Model performance

In the validation cohort, the ResNet101-based deep learning model demonstrated a superior AUC of 0.880 (95% CI 0.780–0.979), indicating its high effectiveness in distinguishing between positive and negative STAS cases, as shown in Figs. 7, 8, Grad-CAM visualizations were used to interpret the model's prediction process, highlighting regions of interest within tumor boundaries. Supplementary Material 2A displays two

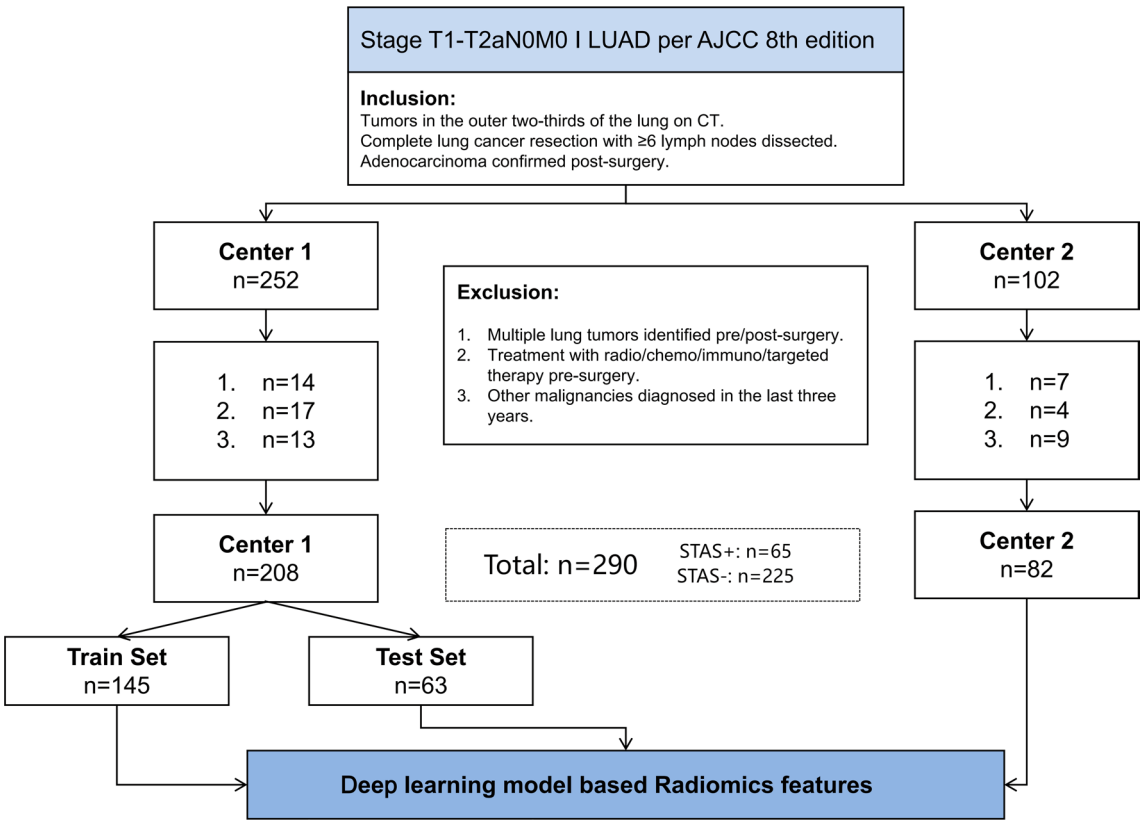

Fig. 1 Screening flowchart for enrolled patients

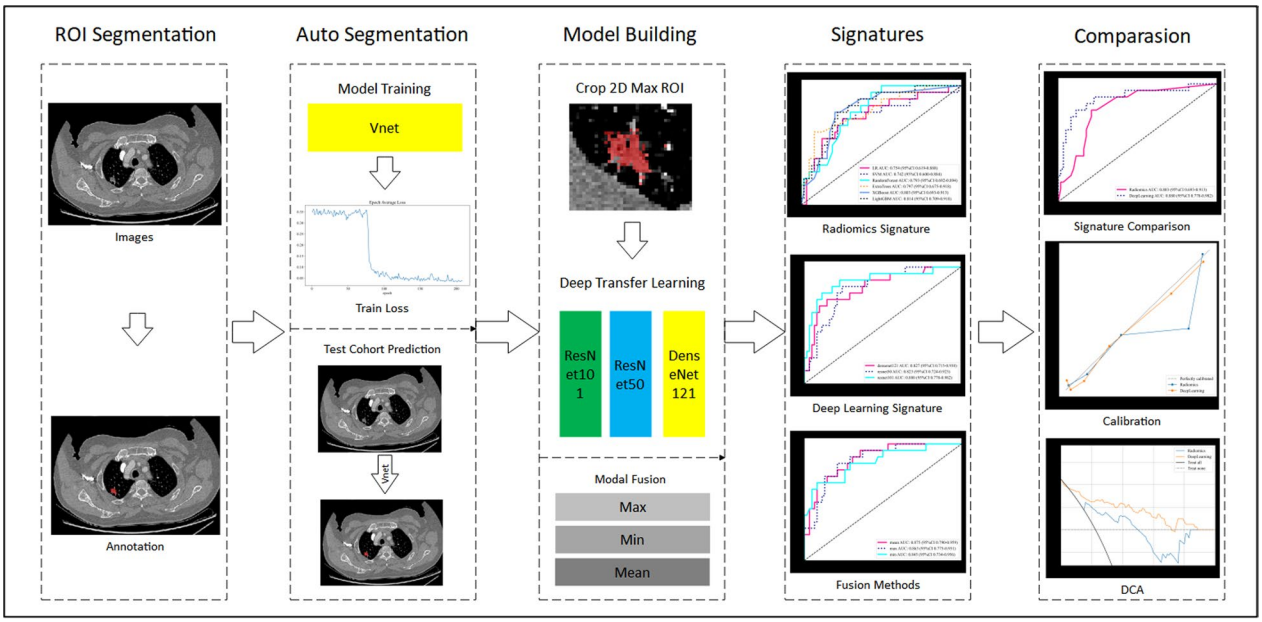

Fig. 2 Workflow of this study

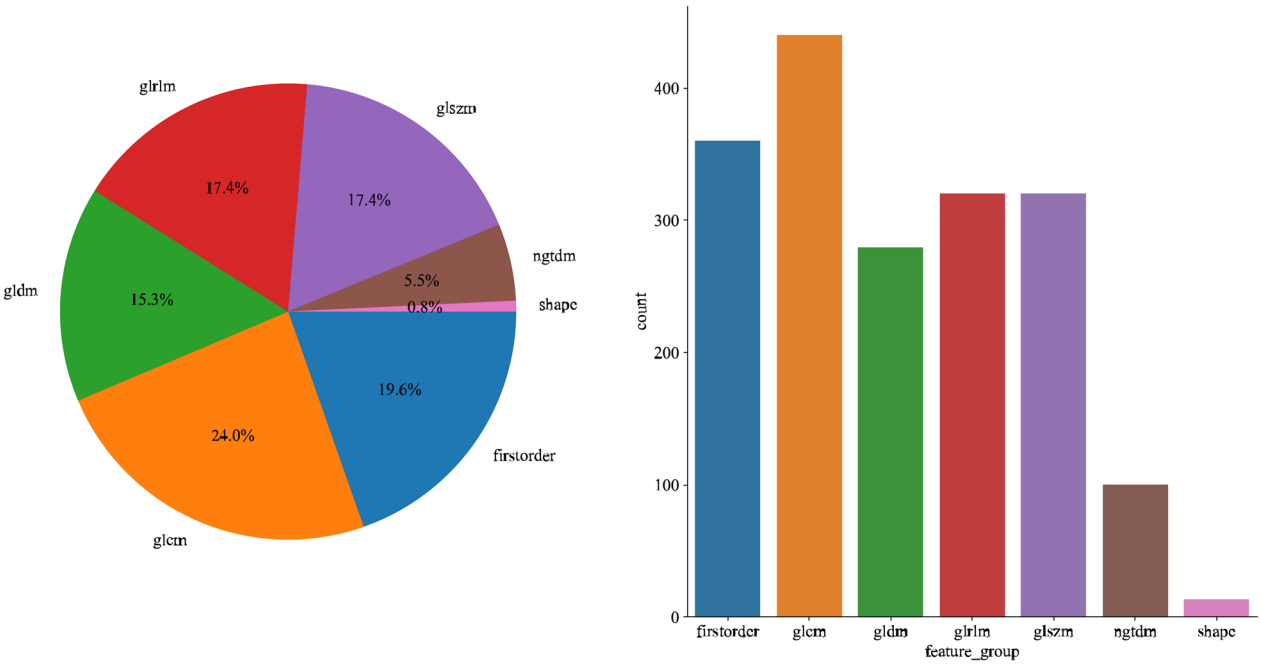

Fig. 3 Number and ratio of handcrafted features

representative cases with their corresponding Grad-CAM visualizations, illustrating the areas most influential in the model's decision-making process. These results underscore the strategic advantage of deep learning models in automatic feature extraction and prediction accuracy.

Signature performance and comparison

A comprehensive analysis of the model performance across different cohorts demonstrated the consistent and superior performance of the DL model over the traditional radiomics model. In the test cohort, the DL model achieved an AUC of 0.880, significantly outperforming

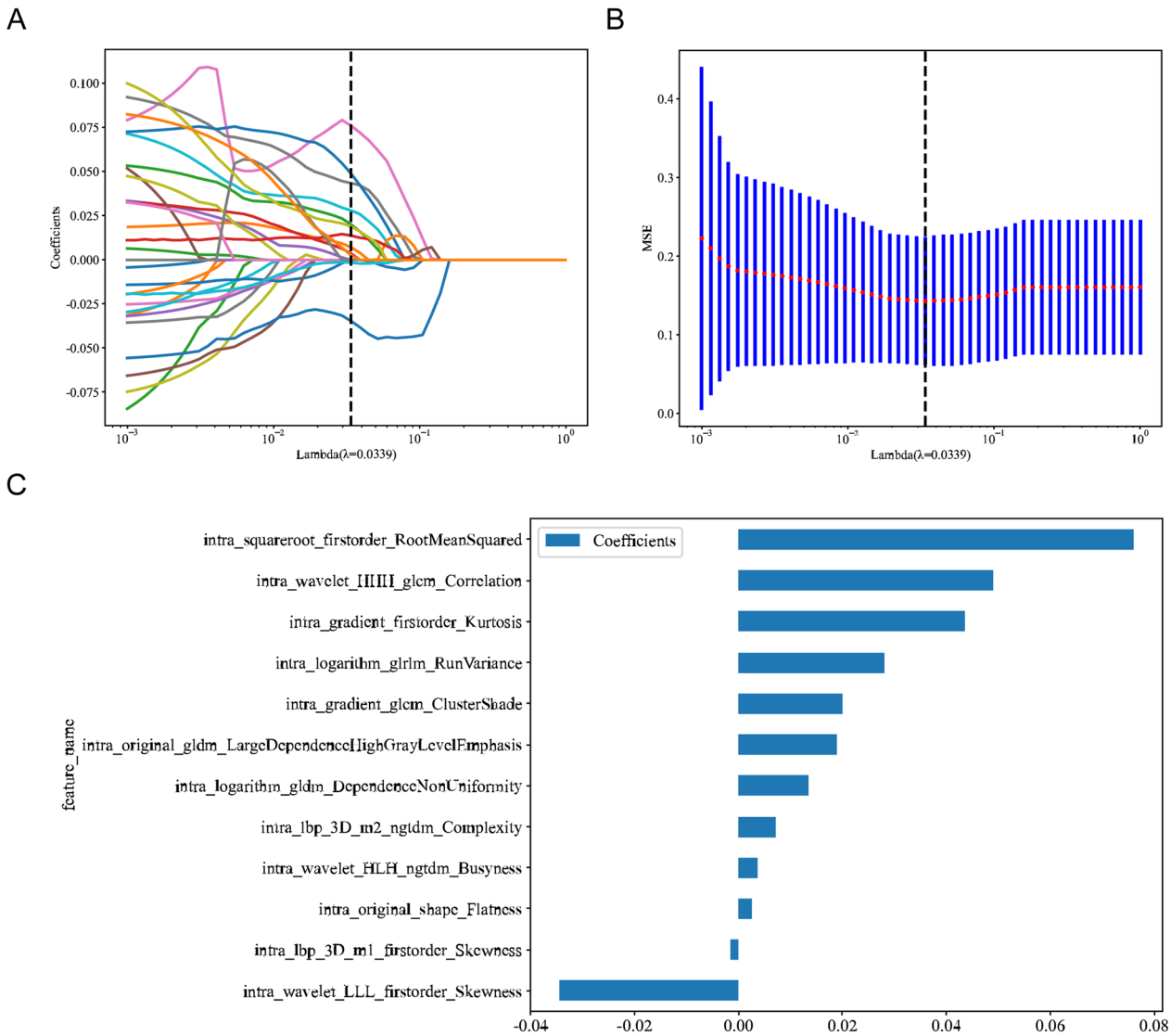

Fig. 4 Coefficients of tenfold cross validation (**A**), MSE of tenfold cross validation (**B**), The histogram of the Rad-score based on the selected features (**C**)

the radiomics model's AUC of 0.803. Similarly, in the validation cohort, the DL signature recorded an AUC of 0.880, compared to 0.833 for the radiomics signature. These results highlight the robustness and generalizability of DL models in STAS prediction, particularly when coupled with automated segmentation techniques. Table 1 and Fig. 9 provide a detailed side-by-side comparison of performance metrics.

Segmentation results

The VNet model for ROI segmentation demonstrated robust performance across training and validation cohorts. The Dice similarity coefficient (DSC) achieved was 0.818 and 0.817, and the Intersection over Union (IoU) was 0.759 and 0.761 for the training and validation cohorts, respectively. Supplementary Material 1A. Table 1 summarizes the segmentation performance metrics. Supplementary Material 1A. Fig. 2 illustrates the segmentation results for representative cases, including comparisons with ground truth annotations. Differences between predicted and ground truth segmentations were minimal, as depicted in the "Diff" column, further validating the segmentation accuracy.

Fusion approaches

To further enhance the prediction accuracy, three fusion approaches (mean, max, min) were employed for combining outputs from multiple models. Fusion

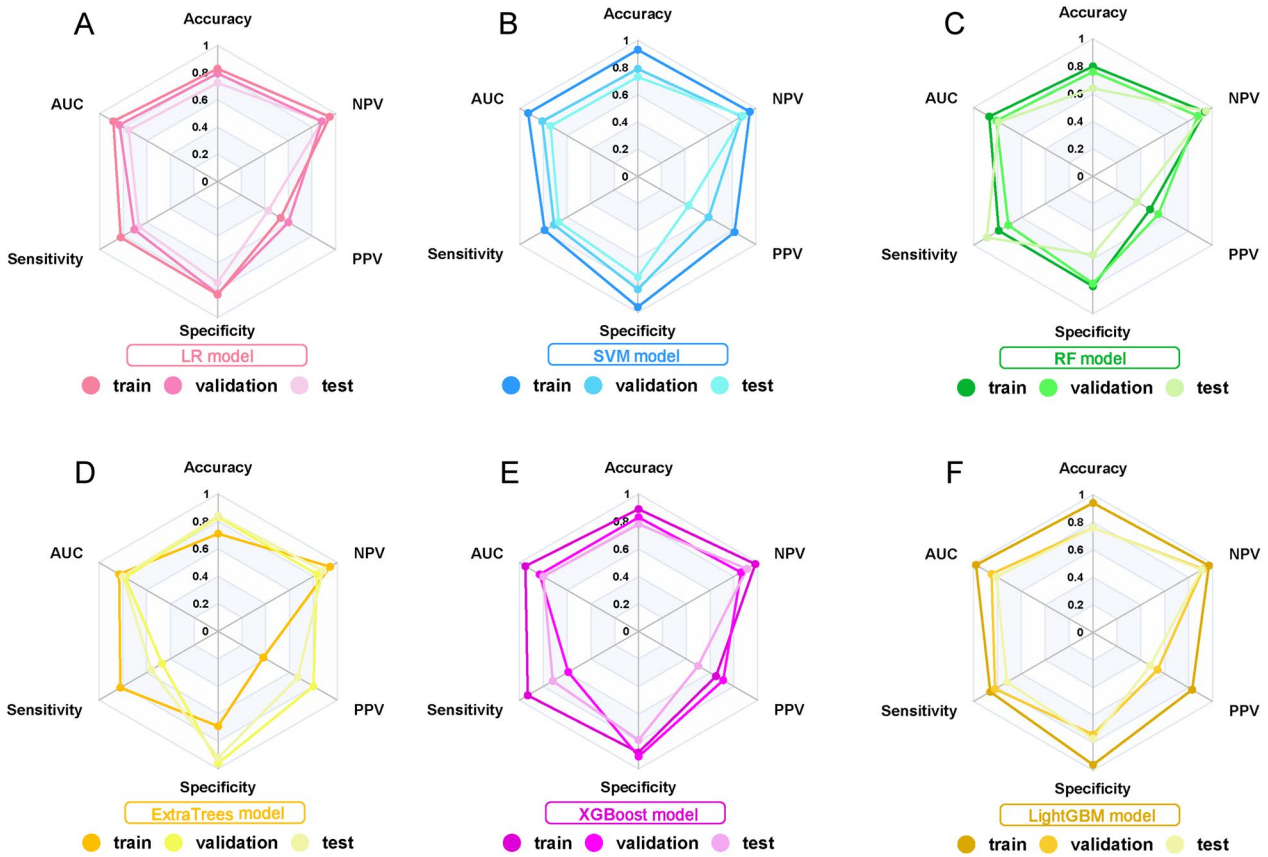

Fig. 5 Metric results for Machine Learning Radiomics Signature. (**A** LR model; **B** SVM model; **C** RF model; **D** ExtraTrees model; **E** XGBoost model; **F** LightGBM model)

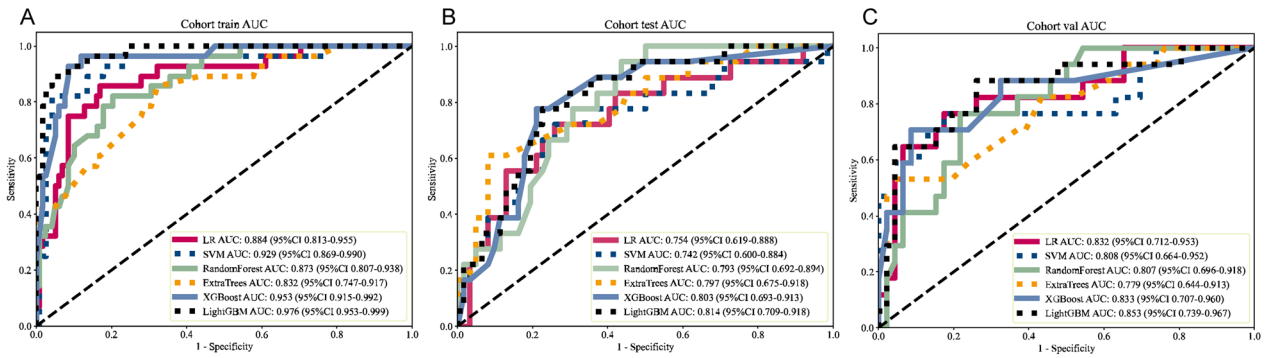

Fig. 6 ROC results for Radiomics Signature of different Machine Learning model. (**A** Cohort train AUC; **B** Cohort test AUC; **C** Cohort validation AUC)

methods showed improvements in AUC values within the training cohort, with the mean fusion method achieving the highest AUC of 0.969. However, the impact of fusion was less pronounced in the validation and test sets, with AUC values of 0.875 and 0.875 for the mean fusion method, respectively. Supplementary Material 2B. Table 1Data for Research Analysis. This file includes and Fig. 2 detail the performance metrics for each fusion approach. The results indicate that while fusion improves performance in controlled conditions, its generalizability to unseen datasets may be limited.

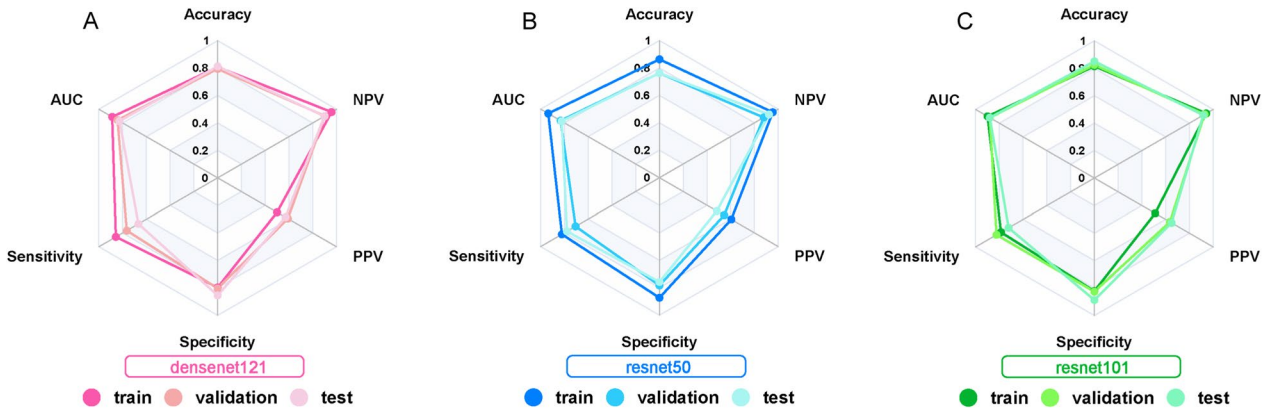

Fig. 7 Metric results for Deep Learning Radiomics Signature (**A** densenet121 model; **B** resnet50 model; **C** resnet101 model)

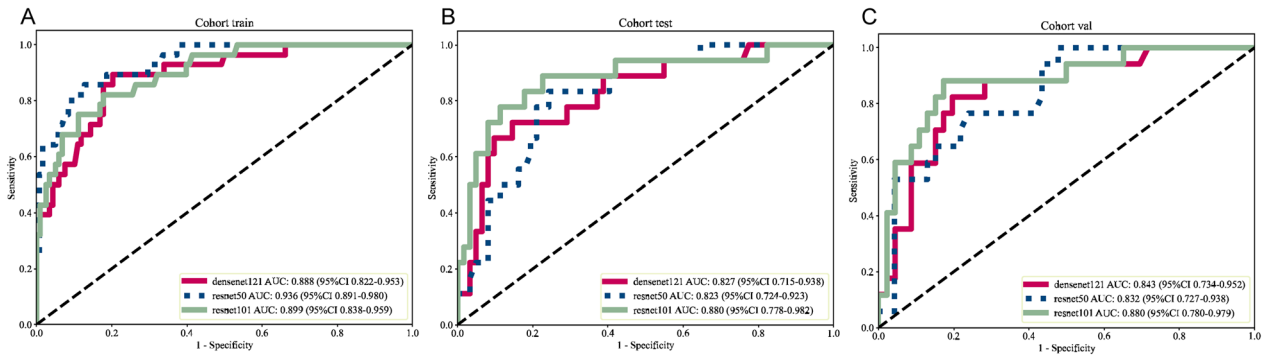

Fig. 8 ROC results for Deep Learning Signature of different model (**A** Cohort train AUC; **B** Cohort test AUC; **C** Cohort validation AUC)

Table 1 Metrics on different signature

Cohort	Signature	ACC	AUC	95% CI	SEN	SPE	PPV	NPV
Train	Radiomics	0.89	0.953	0.915–0.992	0.929	0.881	0.65	0.981
	DL	0.815	0.899	0.838–0.959	0.786	0.822	0.512	0.942
Val	Radiomics	0.825	0.833	0.707–0.960	0.588	0.913	0.714	0.857
	DL	0.825	0.88	0.780–0.979	0.824	0.826	0.636	0.927
Test	Radiomics	0.775	0.803	0.693–0.913	0.722	0.79	0.5	0.907
	DL	0.85	0.88	0.778–0.982	0.722	0.887	0.65	0.917

DL: DeepLearning, ACC: Accuracy, SEN: Sensitivity, SPE: Specificity
PPV: Positive Predictive Value, NPV: Negative Predictive Value

Model calibration and clinical utility

Calibration analysis, assessed using the Hosmer–Lemeshow (HL) test, demonstrated excellent alignment between predicted probabilities and observed outcomes for the DL model. HL test statistics were 0.828, 0.911, and 0.852 for the training, validation, and test cohorts, respectively, indicating the model's superior ability to reflect true event probabilities across cohorts. The calibration curves, shown in Fig. 10, further validate the reliability of the DL model in clinical scenarios.

To assess the clinical utility of the developed models, decision curve analysis (DCA) was conducted. As shown in Fig. 11, the fusion model, which integrates predictions from both radiomics and DL models, provided significant net benefit across a wide range of threshold probabilities. These findings underscore the practical utility of the fusion model in guiding clinical decision-making for STAS assessment, offering an effective balance of predictive accuracy and actionable outcomes.

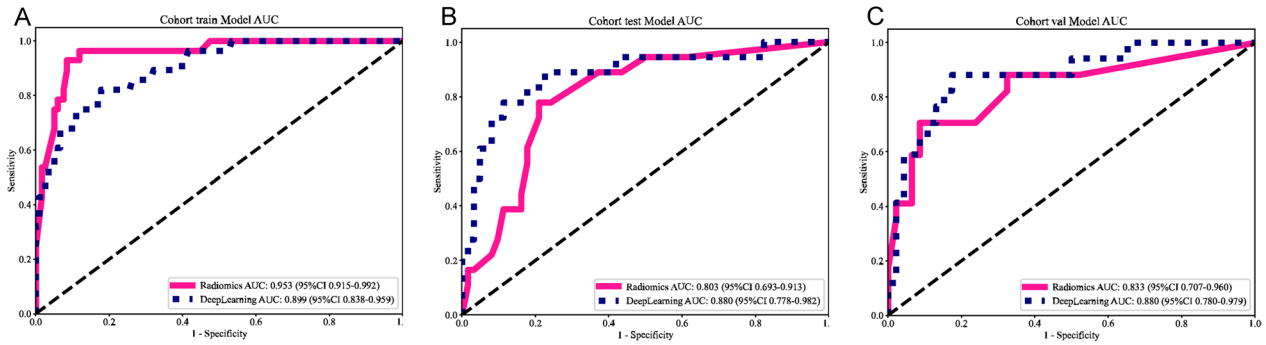

Fig. 9 Illustrates the ROC for different signatures across various cohorts, offering a visual comparison of their diagnostic abilities (**A** Cohort train AUC; **B** Cohort test AUC; **C** Cohort validation AUC)

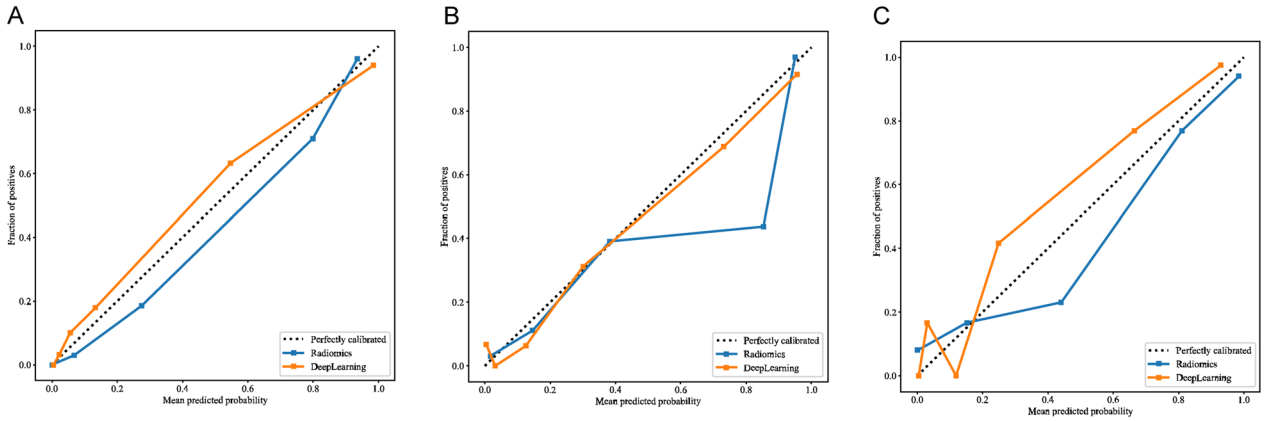

Fig. 10 Displays the calibration curves for different signatures in the various cohorts. These curves are instrumental in understanding how well the predicted probabilities of the models match the actual outcomes (**A** Cohort train calibration curve; **B** Cohort test calibration curve; **C** Cohort validation calibration curve)

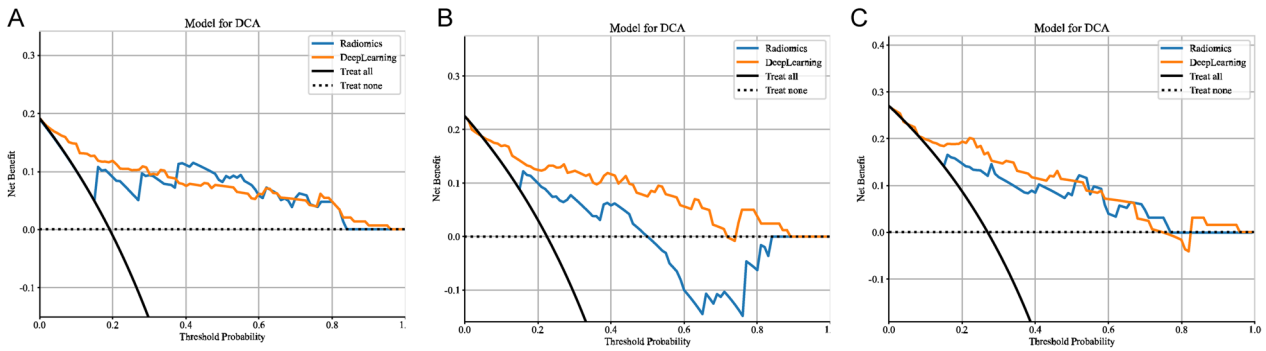

Fig. 11 Different signatures' DCA on various cohorts (**A** Cohort train DCA; **B** Cohort test DCA; **C** Cohort validation DCA)

Discussion

In this study, we developed and evaluated a deep learning (DL) model based on automatic segmentation for predicting STAS in peripheral stage I LUAD. Our DL model achieved an AUC of 0.880 (95% CI 0.780–0.979) in the validation and test datasets, outperforming the radiomics model, which achieved an AUC of 0.833 (95% CI 0.707–0.960). By relying solely on imaging data, we mitigated biases associated with subjective markers or incomplete clinical variables. The integration of automated segmentation using the VNet framework further enhanced workflow efficiency and reproducibility. Moreover,

Grad-CAM visualizations enabled us to identify tumor regions most critical for predictions, improving transparency and interpretability for clinical applications.

We recognize that accurate preoperative prediction of STAS is critical for thoracic surgeons to optimize surgical strategies, especially given the strong association between STAS and poor prognosis in early-stage LUAD. Studies have shown that STAS-positive patients undergoing sublobar resection are at a higher risk of local recurrence and reduced survival compared to those undergoing lobectomy [19, 20]. However, intraoperative frozen section (FS) pathology, the current standard for diagnosing STAS, has limited sensitivity (44%) and moderate accuracy (71%) [21]. These limitations motivated us to explore more reliable, non-invasive methods to predict STAS.

Radiomics has shown promise in addressing this challenge by enabling quantitative, non-invasive STAS prediction from preoperative imaging. Traditional imaging features, such as the consolidation-to-tumor ratio (CTR), pleural indentation, and vascular cancer embolus, have been associated with STAS [22, 23]. However, we recognize that these features are often limited by variability in measurement and subjective interpretation. In contrast, quantitative radiomics extracts a wide range of imaging features and has demonstrated superior predictive performance. For example, Jiang et al. [24] reported an AUC of 0.754 using a random forest model incorporating 12 radiomics features and age, while Liao et al. [25] achieved an AUC of 0.87 (95% CI 0.82–0.92) with a model combining 18 radiomics features and two clinical characteristics. Notably, radiomics models that incorporate peritumoral features tend to perform better than those relying solely on intratumoral characteristics [26]. These findings underscore the importance of incorporating spatially relevant features into STAS prediction.

While radiomics is effective, we acknowledge its limitations, particularly its reliance on manual feature extraction and sensitivity to dataset-specific characteristics, which can hinder generalizability. To overcome these challenges, we leveraged a DL-based approach that incorporates automated segmentation and feature extraction. Using the VNet-based segmentation model, we achieved accuracy comparable to manual delineation while significantly improving workflow efficiency and reproducibility. Furthermore, our DL model's reliance solely on imaging data avoids biases introduced by incomplete clinical information or subjective markers. The Grad-CAM visualizations we generated highlighted tumor regions critical for prediction, enhancing the interpretability and transparency of our DL approach.

To optimize the strengths of both radiomics and DL approaches, we adopted tailored methodologies. Radiomics features were extracted from all slices containing the tumor and aggregated into a case-wise score, effectively capturing tumor heterogeneity. On the other hand, our DL approach focused on the slice with the largest region of interest (ROI), prioritizing computational efficiency while maintaining robustness. These strategies illustrate the complementary strengths of the two approaches: radiomics excels at capturing spatial variability across the tumor, while DL benefits from automated and reproducible feature extraction.

We also explored fusion strategies (mean, max, and min) to improve the DL model's performance by integrating predictions from multiple models. Among these, the mean fusion approach achieved the highest AUC in the training dataset, though its impact on the validation and test datasets was less pronounced. This may reflect the averaging effect, which mitigates outliers but reduces the influence of stronger-performing models. In future studies, we plan to investigate advanced ensemble techniques, such as weighted fusion or attention-based strategies, to better harness the complementary strengths of different models.

When comparing our results to previous studies, such as Wang et al. [27], who reported an AUC of 0.933 for a DL model across LUAD stages I–IV, and Lin et al. [28], who achieved an AUC of 0.82 in a broader cohort (MIA–IIIA), we focused specifically on peripheral stage I LUAD. This narrower scope provides valuable insights into early-stage disease management but may limit the generalizability of our findings to other LUAD stages. We believe that expanding the study population to include a broader range of LUAD stages is a critical next step to validate the robustness of our model across diverse clinical scenarios.

Despite these promising findings, we acknowledge several limitations. First, the dataset size (n = 290) and class imbalance (22% STAS-positive cases) may have influenced model training and evaluation, particularly for metrics like positive predictive value (PPV). While we applied SMOTE to improve PPV for the radiomics model, its application in the DL framework yielded minimal benefits. Future studies should explore advanced balancing techniques, such as cost-sensitive learning or domain-specific data augmentation, to address this issue. Second, the retrospective design and reliance on single-center data for training limit the generalizability of our findings. Although we performed external validation using data from another center, larger multi-center datasets are necessary to ensure robustness and mitigate potential biases. Third, we did not include a direct comparison of DL models with and without segmentation. Such a comparison would provide valuable insights into the incremental value of automated segmentation in

the predictive pipeline, and we plan to prioritize this in future research.

Looking ahead, we believe integrating DL models with multi-modal data, including genomic, histopathological, and tumor microenvironment features, could further enhance predictive accuracy. Combining handcrafted radiomics features with DL-extracted features may also improve performance by leveraging complementary strengths. Prospective studies will be essential to evaluate the real-world performance and clinical utility of these models. Additionally, refining interpretability tools, such as Grad-CAM, will help build clinician trust and facilitate the seamless integration of AI into routine workflows.

Conclusion

We demonstrated that a DL model based on automatic segmentation provides reliable and consistent performance in predicting STAS for peripheral stage I LUAD, outperforming conventional radiomics models. By minimizing manual variability and enhancing reproducibility, our proposed DL approach represents a significant advancement in integrating AI into thoracic oncology workflows. However, addressing dataset limitations and validating the model through multi-center studies will be essential to ensure broader applicability and clinical impact.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12931-025-03174-0.

Supplementary materials 1. Training ROI Auto Segmentation and Model Fusion Details. This file contains detailed information regarding the automatic segmentation model's training process, evaluation metrics, and visualization results. Additionally, it covers Grad-CAM visualizations and the post-fusion techniques applied to multiple models for STAS prediction. The performance of these models is assessed with metrics such as accuracy, AUC, sensitivity, specificity, and others

Supplementary materials 2. Raw Data for Research Analysis. This file includes all the original data used for the analysis in this study. It contains the raw datasets that were processed and analyzed to support our findings

Acknowledgements

Guarantor: The scientifc guarantor of this publication is Xiao-Feng Li. Statistics and Biometry: No complex statistical methods were necessary for this paper.

Author contributions

Conception and design:Xiao-Feng Li, Hui Qian; Administrative support:Xiao-Feng Li, Ping Gong, Yu-Feng Wang; Provision of study materials or patients:Chao Jia, Xiu-Qing Xue; Collection and assembly of data:Hong-Ying Zhao, Cong Liu; Data analysis and interpretation:Cong Liu, Xiao-Feng Li; Manuscript writing: All authors; Final approval of manuscript: All authors.

Funding

This work was supported by Xuzhou Science and Technology Bureau Project [grant number KC23229]; The study was supported by Clinical medicine science and technology development foundation of Jiangsu University, China (Pro. No. JLY2021082).

Data availability

The data and and analysis code used in the current study are available in the open source website github (https://github.com/liucong1994/model-data.git).

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013), The ethics committee of Xuzhou Cancer Hospital approved the study protocol (2023–02-027-K01); This work was supported by the Natural Science Foundation of China (No.82001987); This work was supported by the Key Project of Yancheng Municipal Health Commission (YK2023007).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

Informed consent

Due to the retrospective nature of our study, informed consent from patients was waived.

Author details

[1]Department of Minimally Invasive Oncology, Xuzhou New Health Geriatric Hospital, Xuzhou, People's Republic of China. [2]Departments of Nuclear Medicine, The Xuzhou Hospital Affiliated to Jiangsu University, Xuzhou Cancer Hospital, Xuzhou, People's Republic of China. [3]School of Medical Imaging, Xuzhou Medical University, Xuzhou, People's Republic of China. [4]Department of Nuclear Medicine, The First People's Hospital of Yancheng, Yancheng, People's Republic of China. [5]Department of Radiotherapy, The Xuzhou Hospital Affiliated to Jiangsu University, Xuzhou Cancer Hospital, Xuzhou, People's Republic of China. [6]Medical College of Jiangsu University, Zhenjiang, People's Republic of China. [7]Department of Radiology, The Xuzhou Hospital Affiliated to Jiangsu University, Xuzhou Cancer Hospital, Xuzhou, People's Republic of China.

Received: 19 July 2024 Accepted: 28 February 2025
Published online: 08 March 2025

References

1. Travis WD, Brambilla E, Nicholson AG, et al. WHO panel. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol. 2015;10(9):1243–60.
2. Ren Y, Xie H, Dai C, et al. Prognostic impact of tumor spread through air spaces in sublobar resection for 1A lung adenocarcinoma patients. Ann Surg Oncol. 2019;26:1901–8.
3. Eguchi T, Kameda K, Lu S, et al. Lobectomy is associated with better outcomes than sublobar resection in spread through air spaces (STAS)-positive T1 lung adenocarcinoma: a propensity score matched analysis. J Thorac Oncol. 2019;14:87–98.
4. Suzuki K, Saji H, Aokage K, et al. Comparison of pulmonary segmentectomy and lobectomy: safety results of a randomized trial. J Thorac Cardiovasc Surg. 2019;158(3):895–907.
5. Chen D, Mao Y, Wen J, et al. Tumor spread through air spaces in non-small cell lung cancer: a systematic review and meta-analysis. Ann Thorac Surg. 2019;108(3):945–54.
6. Kadota K, Nitadori JI, Sima CS, et al. Tumor spread through air spaces is an important pattern of invasion and impacts the frequency and location of recurrences after limited resection for small stage I lung adenocarcinomas. J Thorac Oncol. 2015;10(5):806–14.
7. Walts AE, Marchevsky AM. Current evidence does not warrant frozen section evaluation for the presence of tumor spread through alveolar spaces. Arch Pathol Lab Med. 2018;142:59–63.

8. Villalba JA, Shih AR, Sayo TMS, et al. Accuracy and reproducibility of intraoperative assessment on tumor spread through air spaces in stage 1 lung adenocarcinomas. J Thorac Oncol. 2021;16(4):619–29.
9. Zhou F, Villalba JA, Sayo TMS, et al. Assessment of the feasibility of frozen sections for the detection of spread through air spaces (STAS) in pulmonary adenocarcinoma. Mod Pathol. 2022;35(2):210–7.
10. Zhang Y, Liao Q, Ding L, et al. Bridging 2D and 3D segmentation networks for computation-efficient volumetric medical image segmentation: an empirical study of 2.5D solutions. Comput Med Imaging Graph. 2022;99:102088.
11. Xie H, Chen Z, Deng J, et al. Automatic segmentation of the gross target volume in radiotherapy for lung cancer using transresSEUnet 2.5D Network. J Transl Med. 2022;20(1):524.
12. Kalpathy-Cramer J, Zhao B, Goldgof D, et al. A comparison of lung nodule segmentation algorithms: methods and results from a multi-institutional study. J Digit Imaging. 2016;29(4):476–87.
13. Sumida I, Magome T, Kitamori H, et al. Deep convolutional neural network for reduction of contrast-enhanced region on CT images. J Radiat Res. 2019;60(5):586–94.
14. Brosch T, Tang LY, Youngjin Y, et al. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. IEEE Trans Med Imaging. 2016;35(5):1229–39.
15. Detterbeck FC, Boffa DJ, Kim AW, et al. The Eighth Edition Lung Cancer Stage Classification. Chest. 2017;151(1):193–203.
16. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:5527.
17. Kocak B, Baessler B, Bakas S, et al. CheckList for EvaluAtion of radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. Insights Imag. 2023;14(1):75.
18. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillon-Robin JC, Pieper S, Aerts HJWL. Computational radiomics system to decode the radiographic phenotype. Can Res. 2017;77(21):e104–7.
19. Liu A, Hou F, Qin Y, et al. Predictive value of a prognostic model based on pathologic features in lung invasive adenocarcinoma. Lung Cancer. 2019;131:14–22. https://doi.org/10.1016/j.lungcan.2019.03.002.
20. Vaghjiani RG, Takahashi Y, Eguchi T, et al. Tumor spread through air spaces is a predictor of occult lymph node metastasis in clinical stage IA lung adenocarcinoma. J Thorac Oncol. 2020;15(5):792–802. https://doi.org/10.1016/j.jtho.2020.01.008.
21. Villalba JA, Shih AR, Sayo TMS, et al. Accuracy and reproducibility of intraoperative assessment on tumor spread through air spaces in stage 1 lung adenocarcinomas. J Thorac Oncol. 2021;16(4):619–29. https://doi.org/10.1016/j.jtho.2020.12.005.
22. Jia C, Jiang HC, Liu C, et al. The correlation between tumor radiological features and spread through air spaces in peripheral stage IA lung adenocarcinoma: a propensity score-matched analysis. J Cardiothorac Surg. 2024;19(1):19. https://doi.org/10.1186/s13019-024-02498-0.
23. Wang J, Yao Y, Tang D, Gao W. An individualized nomogram for predicting and validating spread through air space (STAS) in surgically resected lung adenocarcinoma: a single center retrospective analysis. J Cardiothorac Surg. 2023;18(1):337. https://doi.org/10.1186/s13019-023-02458-0.
24. Jiang C, Luo Y, Yuan J, et al. CT-based radiomics and machine learning to predict spread through air space in lung adenocarcinoma. Eur Radiol. 2020;30(7):4050–7. https://doi.org/10.1007/s00330-020-06694-z.
25. Liao G, Huang L, Wu S, et al. Preoperative CT-based peritumoral and tumoral radiomic features prediction for tumor spread through air spaces in clinical stage I lung adenocarcinoma. Lung Cancer. 2022;163:87–95. https://doi.org/10.1016/j.lungcan.2021.11.017.
26. Liu C, Wang YF, Wang P, et al. Predictive value of multiple imaging predictive models for spread through air spaces of lung adenocarcinoma: a systematic review and network meta-analysis. Oncol Lett. 2024;27(3):122. https://doi.org/10.3892/ol.2024.14255.
27. Wang S, Liu X, Jiang C, et al. CT-based super-resolution deep learning models with attention mechanisms for predicting spread through air spaces of solid or part-solid lung adenocarcinoma. Acad Radiol. 2024. https://doi.org/10.1016/j.acra.2023.12.034.
28. Lin MW, Chen LW, Yang SM, et al. CT-based deep-learning model for spread-through-air-spaces prediction in ground glass-predominant lung adenocarcinoma. Ann Surg Oncol. 2024;31(3):1536–45. https://doi.org/10.1245/s10434-023-14565-2.

Publisher's Note