npj | precision oncology

Article

Published in partnership with The Hormel Institute, University of Minnesota

https://doi.org/10.1038/s41698-026-01305-4

Integrated predictive model for visceral pleural invasion in small NSCLC with high clinical utility

Shuyi Yang[1,2,5], Ying Wei[3,5], Qingle Wang[1,2], Yaoyao Zhuo[1,2], Shan Yang[1,2], Weiya Shi[4], Yinwen Gan[4], Tingting Cai[4], Yichu He[3], Yi Zhan[4], Haoling Zhang[1,2], Yuxin Shi[2,4], Mengsu Zeng[1,2], Feng Shi[3], Zhong Xue[3] ✉, Zhiyong Zhang[1,2] ✉ & Fei Shan[2,4] ✉

This study aims to develop and validate a multi-feature integrated imaging fusion (MIIF) model, incorporating deep learning, radiomics features, and computed tomography (CT) findings, for identifying visceral pleural invasion (VPI) in small non-small cell lung cancer (NSCLC). This multi-center retrospective analysis included 2822 small NSCLCs. These were divided into four datasets (training, validation, internal/external test). The MIIF model's diagnostic performance was compared against the assessments of six radiologists. Additionally, we evaluated the clinical utility of the MIIF model by comparing the diagnostic performance of radiologists, with/without the aid of the model. The MIIF model yielded AUCs of 0.869/0.785 in the internal/external test sets, respectively, which were comparable to the radiologists' ($P > 0.05$). With MIIF assistance, radiologists' accuracy and specificity increased to 0.845/0.828 and 0.836/0.841 in the internal/external test sets ($P < 0.001$). The MIIF model shows enhanced accuracy and specificity in detecting VPI in small NSCLC and may improve radiologist' diagnostic performance.

Lung cancer is the leading cause of cancer-related death worldwide, and non-small cell lung cancer (NSCLC) is one of the most common types[1]. Among NSCLC patients, lung adenocarcinoma is frequently observed as a histological subtype. Visceral pleural invasion (VPI), which refers to the tumor's penetration beyond the elastic layer of the visceral pleura and encompasses levels Pl1 and Pl2, is critical in the staging of the tumor-node-metastasis (TNM) system and treatment planning. Specifically, VPI escalates the TNM T stage to T2a when the tumor's solid component measures up to 30 mm, significantly affecting treatment strategies[2–4]. The survival outcomes vary distinctly across stages, for instance, the estimated 5-year overall survival rate for clinical stage IA lung cancer stands at 82%, whereas it drops to 69% for stage IB (T2aN0M0) disease[2]. Results from a secondary analysis of a randomized clinical trial indicated that patients with small NSCLC (≤20 mm) exhibiting VPI experienced poorer disease-free and recurrence-free survival, as well as a higher incidence of local and distant disease recurrence[5]. Additionally, VPI was associated with nodal involvement and skip N2 metastases in small NSCLC, which complicates disease management[3,4]. Thus, precise preoperative assessment of VPI is essential for determining the optimal surgical approach, particularly regarding the timing of surgery and the extent of lymphadenectomy.

However, accurately diagnosing VPI preoperatively poses a significant challenge, particularly in cases of small NSCLC (solid component ≤ 30 mm). In recent years, researchers have investigated specific computed tomography (CT) findings to predict VPI, yielding mixed outcomes[6–9]. Sun Q et al.[6] initially identified a CT manifestation known as the jellyfish sign, which reliably predicted VPI, showing an odds ratio (OR) of 21.6 ($P < 0.001$). Onoda H et al.[7] examined tumors that do not appear touching the pleural surface and introduced the bridge tag sign to potentially enhance VPI prediction accuracy. Yang S et al.[8] first proposed a parameter, called pleural indentation fraction (PIF), which defined as the ratio of pleural shift distance to the projected length of involved pleura, for quantifying pleural shifts[8]. However, the validity of these CT features as substitutes for clinical T2 descriptors remains controversial and warrants further confirmation. These limitations underscore the necessity for more accurate and thorough methods for VPI predicion.

In the realm of medical imaging analysis, Artificial intelligence (AI) has shown considerable promise, integrating techniques such as radiomics and

[1]Department of Radiology, Zhongshan Hospital, Fudan University, Shanghai, China. [2]Shanghai Institute of Medical Imaging, Shanghai, China. [3]Department of Research and Development, United Imaging Intelligence, Shanghai, China. [4]Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, Shanghai, China. [5]These authors contributed equally: Shuyi Yang, Ying Wei. ✉e-mail: zhong.xue@uii-ai.com; zhyzhang@fudan.edu.cn; shanfei_2901@163.com

deep learning (DL). Radiomics involves extracting quantitative attributes from images, capturing minute tissue details that might escape visual detection[10]. Nonetheless, the utility of radiomics is somewhat limited by its dependence on predefined algorithms for feature extraction.

Conversely, DL technology enables in-depth exploration of high-dimensional quantification of radiological images and automatically learns hierarchical features from raw image data. DL is proficient in capturing complex patterns and has demonstrated superior performance in various medical imaging applications[11,12]. Despite some studies applying DL to predict VPI status, the reported area under the receiver operating characteristic curve (AUC) values were comparable to those from radiologists' evaluation or clinical models[13–15]. However, the efficiency of these DL models is not yet satisfactory, and their clinical benefits needs further assessment[11].

Recognizing the complementary strengths of these methodologies, recent studies have investigated the integration of radiomics and DL features[16,17]. This synergy promises to exploit both the interpretable, biologically relevant features of radiomics and the high-level abstract features acquired by deep neural networks. In the realm of VPI status, such an integrated approach could provide a more comprehensive characterization of tumor properties, potentially enhancing model performance and robustness.

Therefore, our study aimed to utilize the complementary nature of handcrafted radiomics and DL by developing a multi-feature integrated imaging fusion (MIIF) model, which integrates general CT findings, radiomics features, and deep imaging features. This model is designed for risk assessment of VPI status in patients with small NSCLC (solid component ≤ 30 mm) on preoperative CT images. Furthermore, we evaluated the model's utility in clinical practice using a paired design to compare the diagnostic performance of radiologists, both without and with the assistance of our proposed model. We also investigated the CT findings for predicting VPI.

Results

Basic characteristics of patients

A total of 2698 patients with 2822 pathologically confirmed NSCLCs were enrolled in this study (Fig. 1). The characteristics of the patients were detailed in Table 1. VPI was present in 20.61% (408/1980) of lesions in the training set and 22.64% (91/402) in the validation set. In the internal and external test sets, VPI was observed in 9.75% (27/277) and 15.34% (25/163) of lesions, respectively.

Diagnostic performance of the proposed model

The diagnostic performances of the DL and MIIF models for predicting VPI was shown in Table 2 and Supplementary Table 1. The MIIF model, which incorporated 42 significant features identified by the least absolute shrinkage and selection operator (LASSO) algorithm, included 29 deep imaging features, 9 radiomics features and 4 CT findings (Fig. 2). This model demonstrated improved performance in the validation, internal and external test sets. It yielded an AUC of 0.978 (95% CI: 0.973–0.984) and an accuracy of 0.922 in the training set, along with an AUC of 0.864 (95% CI: 0.828–0.899) and an accuracy of 0.821 in the validation set. In the internal and external test sets, the MIIF model achieved AUCs of 0.869 (95% CI: 0.817–0.921) and 0.785 (95% CI: 0.703–0.867), with accuracies of 0.812 and 0.798, respectively, significantly surpassing the DL model, which showed AUCs of 0.794 (95% CI: 0.722–0.865) and 0.679 (95% CI: 0.575–0.782) and accuracies of 0.690 and 0.644, respectively ($P < 0.001$).

In the internal and external test sets, the MIIF model achieved the smaller Brier score of 0.122 and 0.154, compared to DL model (0.191 and 0.226). The decision curves analysis (DCA) curves (bottom row) demonstrate that both models provide a positive net benefit across a wide range of threshold probabilities (≈0.0–0.2) in all test sets, including the low-prevalence internal test (prevalence = 9.75%). The results of calibration curve analyses and DCA across all datasets are shown in Supplementary Fig. 1.

Diagnostic performance using CT alone and CT with MIIF model assistance

The diagnostic performance of six radiologists, with and without MIIF model assistance, is detailed in Table 3. The MIIF model achieved higher AUCs than each radiologists in both internal and external test sets (0.767-0.839 and 0.656-0.724). The average AUCs of all radiologists in the internal/external test sets were slightly higher/lower than those of the MIIF model, without significant statistical differences (0.879 & 0.869/0.739 & 0.785, $P > 0.05$). The MIIF model generally displayed better accuracies ($P = 0.006$/0.001) and specificities (all $P < 0.001$) in both test sets, particularly for junior radiologists ($P = 0.010$/0.003, all $P = 0.002$), although the sensitivities did not show significant differences ($P > 0.05$).

In the internal test set, the average AUC of radiologists improved from 0.879 without MIIF assistance to 0.921 with MIIF assistance ($P = 0.073$). Similarly, in the external test set, the average AUC increased from 0.739 to 0.828 ($P = 0.003$). These improvements were particularly notable for junior radiologists, whose average AUC in the internal test set improved from 0.853 to 0.904 ($P = 0.036$), and in the external test set from 0.731 to 0.824 ($P < 0.001$). Senior radiologists also showed improvements ($P = 0.105$, 0.008). Each radiologist demonstrated higher AUCs with MIIF model assistance ($P < 0.05$) (Fig. 3).

The MIIF model significantly enhanced radiologists' accuracy and specificity across both test sets. In the internal test set, average accuracy improved from 0.736 to 0.845 ($P < 0.001$), and specificity increased from 0.720 to 0.836 ($P < 0.001$). Similar trends were observed in the external test set, where accuracy increased from 0.663 to 0.828 ($P < 0.001$), and specificity improved from 0.674 to 0.841 ($P < 0.001$). Sensitivity showed modest improvements, though these were not statistically significant ($P > 0.05$).

Associations of specific CT findings with VPI

Specific CT findings associated with VPI were summarized in Table 4. In the internal and external test sets, there were 81 ground-glass nodules (GGNs), 246 part-solid nodules (PSNs), and 113 solid nodules (SNs), with VPI present in 11.82% (52/440) of these nodules. No GGNs exhibited VPI, while VPI was observed in 6.91% of PSNs (17/246) and 30.97% of SNs (35/113) (Supplementary Table 2).

Among the PSNs and SNs, there were 160 pleural-attached nodules (24, 15% VPI positive), 194 pleural-tag nodules (23, 11.86% VPI positive), and 5 nodules pushed against the pleura (all VPI positive) (Fig. 4).

Univariable analysis revealed the statistically significant features according to the VPI status. Multivariable logistic regression identified nodule type (OR, 4.86; $P = 0.036$], solid component mean diameter (cut-off value: 15 mm; OR, 4.67; $P = 0.024$) of pleural-attached nodules, and solid component mean diameter (cut-off value: 13 mm; OR, 3.94; $P = 0.026$), and PIF (cut-off value: 0.405; OR, 23.32; $P = 0.003$) of pleural-tag nodules as independent predictors for VPI, with the exception of the jellyfish and bridge tag signs (Supplementary Table 3).

Discussion

In our study, we developed and validated CT-based DL and MIIF models for preoperative VPI prediction in NSCLC, of which the solid component size was 30 mm or smaller. The diagnostic performance of our proposed model was comparable to that of radiologists, exhibiting significantly higher accuracy and specificity. The high clinical utility if the model was demonstrated through a paired design, which showed improved radiologists' performance with MIIF assistance. Additionally, we categorized subpleural NSCLCs into three groups and identified nodule type, solid component mean diameter, and PIF value as independent predictors of VPI.

These findings have significant clinical implications, especially in guiding treatment decisions. For instance, T1-sized tumors with VPI (stage IB) require more extensive lymph node (LN) dissection than those without VPI (stage IA)[18]. The MIIF model's capacity to enhance diagnostic accuracy, particularly for junior radiologists, can standardize VPI assessments and guide surgical planning. In cases where VPI is highly suspected, prompt surgical intervention is crucial. Integrating the MIIF model into

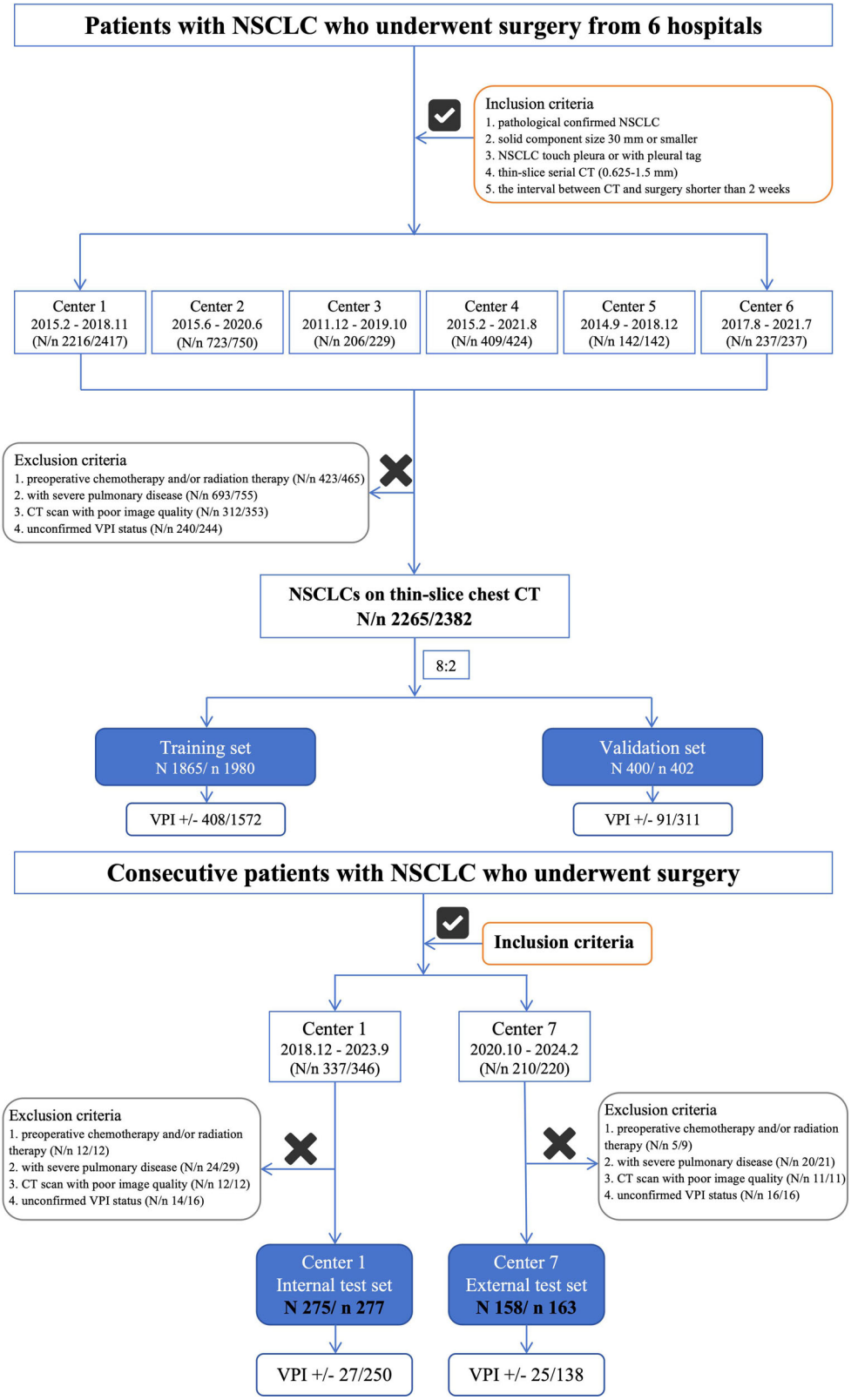

Fig. 1 | Patient inclusion and exclusion criteria. NSCLC non-small cell lung cancer, VPI visceral pleural invasion.

multidisciplinary discussions could improve treatment strategies, especially for small subpleural NSCLC, where accurate staging is vital for optimizing patient outcomes[19,20].

Previous studies have assessed the role of DL in predicting VPI. Choi H et al.[14] developed a DL model that matched radiologist-level performance, achieving an AUC of 0.75, with the advantage of adjustable sensitivity and specificity to meet clinical needs. Similarly, Lim WH et al.[15] developed a DL model for VPI prediction with an AUC of 0.79, comparable to the pooled AUC of 0.78 reported for radiologists. While these studies highlight the potential of DL in VPI prediction, they also underscore the limitations of current models, indicating the necessity for more advanced approaches.

Table. 1 | Baseline characteristics of the four data sets

Clinical characteristic	Training (*N*/*n*, 1865/1980)	Validation (*N*/*n*, 400/402)	Internal test (*N*/*n*, 275/277)	External test(*N*/*n*, 158/163)	*P* value
Mean age (y) (range)	59 ± 12 (16–87)	60 ± 12(26–85)	58 ± 12(16–84)	62 ± 12(22–83)	**0.002**[a]
Sex					0.962
Male	678 (36.35%)	146 (36.50%)	103 (37.45%)	61 (38.61%)	
Female	1187 (63.65%)	254 (63.50%)	172 (62.55%)	97 (61.39%)	
Mean diameter (mm)	14.93 ± 7.08 (2.92–39.31)	14.67 ± 6.78(2.29–36.99)	14.60 ± 5.34(5.27–28.66)	15.93 ± 5.89(5.51–31.54)	0.193
Nodule type					0.103
GGN	323 (16.31%)	80 (19.90%)	49 (17.69%)	32 (19.63%)	
PSN	1226 (61.92%)	230 (57.21%)	162 (58.48%)	84 (51.53%)	
SN	431 (21.77%)	92 (22.89%)	66 (23.83%)	47 (28.83%)	
Solid component mean diameter (PSN + SN)	14.07 ± 5.75 (3.13–29.94)	13.47 ± 5.56 (4.26–29.43)	12.46 ± 6.35 (2.30–28.66)	14.36 ± 5.97 (2.95–29.50)	**0.002**[a]
Location					0.950
Right upper	603 (30.45%)	131 (32.59%)	90 (32.49%)	47 (28.83%)	
Right middle	209 (10.56%)	37 (9.20%)	31 (11.19%)	15 (9.20%)	
Right lower	373 (18.84%)	77 (19.15%)	46 (16.61%)	33 (20.25%)	
Left upper	573 (28.94%)	120 (29.85%)	77 (27.80%)	52 (31.90%)	
Left lower	222 (11.21%)	37 (9.20%)	33 (11.91%)	16 (9.82%)	
Histology					0.962
MIA	366 (18.48%)	75 (18.66%)	53 (19.13%)	30 (18.40%)	
IAC	1573 (79.44%)	321 (79.85%)	219 (79.06%)	130 (79.75%)	
SCC	11 (0.56%)	3 (0.75%)	2 (0.72%)	2 (1.23%)	
ASC	22 (1.11%)	2 (0.50%)	2 (0.72%)	0 (0)	
LCLC	8 (0.40%)	1 (0.25%)	1 (0.36%)	1 (0.61%)	
VPI status					**<0.001**
positive	408 (20.61%)	91 (22.64%)	27 (9.75%)	25 (15.34%)	
negative	1572 (79.39%)	311 (77.36%)	250 (90.25%)	138 (84.66%)	

GGN ground glass nodule, *PSN* part solid nodule, *SN* solid nodule, *MIA* minimally invasive adenocarcinoma, *IAC* invasive adenocarcinoma, *SCC* squamous cell carcinoma, *ASC* adenosquamous carcinoma, *LCLC* large cell lung cancer, *VPI* visceral pleural invasion.

[a]Mean age: statistic difference between training and external test set, as well as internal and external test set; Solid component mean diameter: statistic difference between training and internal test set, as well as internal and external test set.

Bold: statistically significant data.

In this study, our multi-feature integrated approach significantly enhanced VPI prediction in small NSCLC, achieving excellent performance in the internal test set (AUC = 0.869) and acceptable performance in the external test set (AUC = 0.785). This improvement can be attributed to several key factors in our model construction: 1) The attention mechanism in our DL model focuses on relevant image regions, emulating radiologist assessments; 2) the integration of radiomics features and automatically obtained general CT findings offers a comprehensive depiction of tumor characteristics, bridging computational analysis and radiological expertise; 3) The fusion of multiple features capitalizes on the strengths of each modality, allowing for more refined predictions. This integrated approach not only improves model performance but also enhances its generalizability and clinical interpretability, making it a valuable tool for VPI prediction in small NSCLC.

The incorporation of the MIIF model into radiologists' workflow demonstrated its potential to significantly enhance diagnostic performance. In the internal test cohort, the mean AUC increased from 0.879 to 0.921 ($P = 0.073$), and accuracy improved from 0.736 to 0.845 ($P < 0.001$). These improvements were consistent across multiple observers, demonstrating the model's ability to standardize diagnostic evaluations and reduce variability.

The radiologists in our study demonstrated unbalanced diagnostic performance. For example, a senior radiologist demonstrated high specificity (0.912) but low sensitivity (0.704) in VPI assessment, while junior radiologists tended to show either high sensitivity with low specificity or vice verse[14,15]. This variability suggests that inexperience may lead to either overestimation or underestimation of VPI risk using preoperative CT. Accurate identification of VPI in small NSCLC remains complex and challenging in current clinical practice, even with the advent of specific CT findings, underscoring the task's complexity and the need for more reliable diagnostic tools.

The MIIF model effectively addressed these challenges, particularly benefiting junior radiologists. Almost all radiologists exhibited improved accuracies and specificities with the assistance of the MIIF model. Notably, the model also enhanced diagnostic sensitivity for a senior radiologist who initially had low sensitivity (0.704), potentially assisting surgeons in determining the optimal extent of LN dissection. These findings emphasize the high clinical utility of AI-assisted tools in enhancing diagnostic precision and reducing inter-observer variability.

Our analysis of specific CT findings revealed that SNs exhibited a higher rate of VPI (30.97%) compared to PSNs (6.91%), while no VPI was observed in GGNs. These findings align with previous studies[6,19]. However, a few studies have shown that VPI can be present in GGNs at rates ranging from 4.7% to 17.4%[21,22], which remains controversial in clinical practice. In our study, NSCLC with VPI was larger than that without VPI; the solid component size served as an independent predictor of VPI, with an OR of 4.67/3.94 for pleural-attached/tag nodule.

Table 2 | Diagnostic performance of the established deep learning (DL) model and multimodal integrated imaging fusion (MIIF) model

Dataset	Model	AUC	*P* value	Accuracy	*P* value	Sensitivity	*P* value	Specificity	*P* value	PPV	NPV	F1-Score
Training set	DL model	0.997[0.994,0.999]		0.962		0.995		0.953		0.844	0.999	0.913
	MIIF model	0.978[0.973,0.984]	**<0.001**	0.922	**<0.001**	0.914	**<0.001**	0.924	**<0.001**	0.758	0.976	0.829
Validation set	DL model	0.842[0.803,0.881]		0.801		0.703		0.830		0.547	0.905	0.615
	MIIF model	0.864[0.828,0.899]	**<0.001**	0.821	**0.013**	0.747	0.134	0.842	0.134	0.581	0.919	0.654
Internal test set	DL model	0.794[0.722,0.865]		0.690		0.778		0.680		0.208	0.966	0.328
	MIIF model	0.869[0.817,0.921]	**<0.001**	0.812	**<0.001**	0.704	0.480	0.824	**<0.001**	0.302	0.963	0.422
External test set	DL model	0.679[0.575,0.782]		0.644		0.680		0.638		0.254	0.917	0.370
	MIIF model	0.785[0.703,0.867]	**<0.001**	0.798	**<0.001**	0.640	1.000	0.826	**<0.001**	0.4	0.927	0.492

Values within the brackets were 95% confidence intervals.
P value derived from Delong test for AUC, or McNemar test for accuracy, sensitivity and specificity.
Bold: statistically significant data.

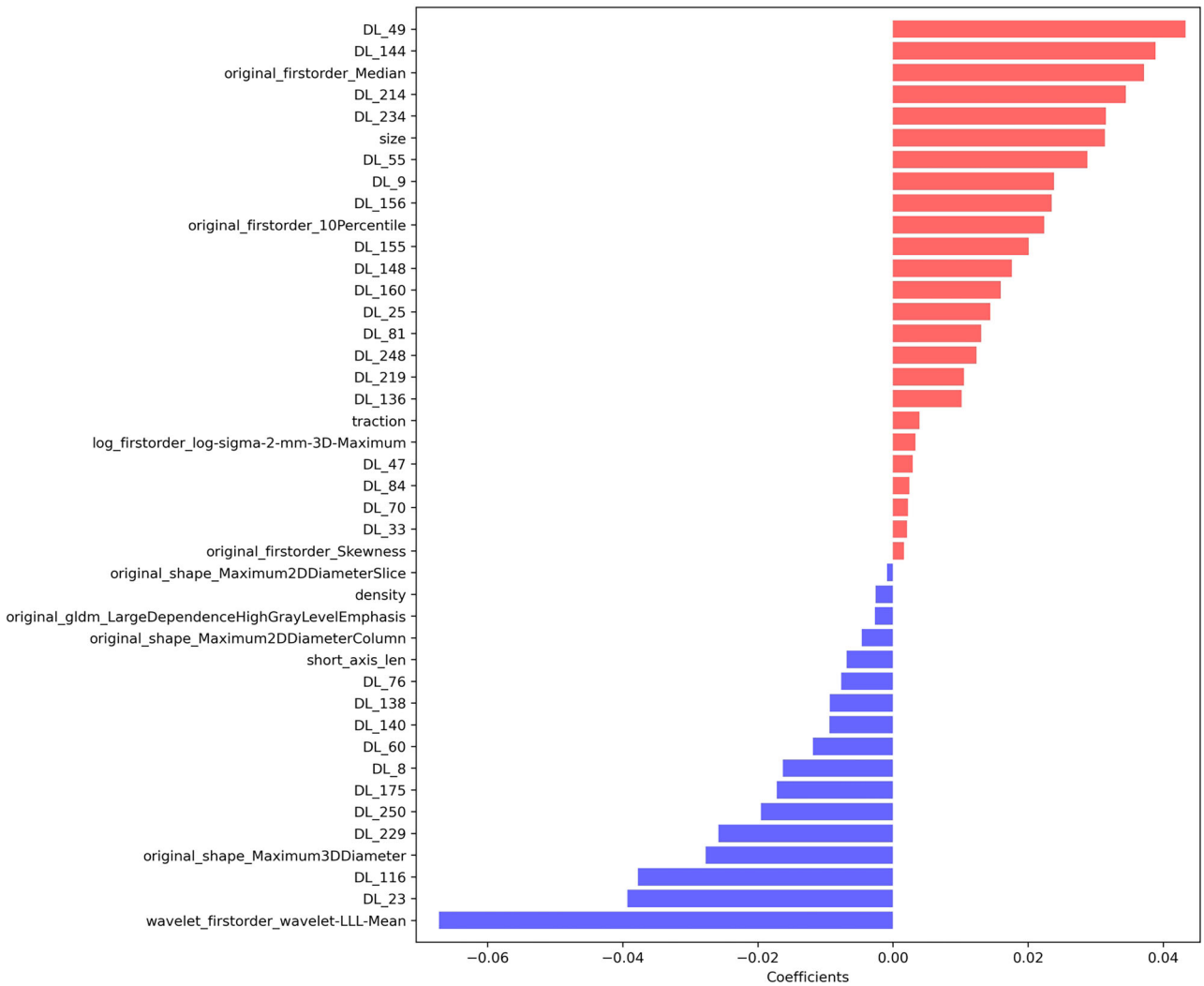

Fig. 2 | Bar plot of significant features associated with VPI status.

The solid component, representing the invasive portion of the tumor[2], is closely associated with increased malignancy risk[23,24], which may clarify this result. NSCLCs pushed against the interlobar pleura in our study suggested tumor penetration through this pleura, with all cases showing VPI, consistent with prior research[8]. These predictors can be readily assessed on preoperative CT scans, serving as valuable tools for clinical decision-making.

Pleural-attached nodules are in direct contact with the pleura, while pleural-tag nodules have one or more linear tag connections[6,7,9,19,25]. Our study showed that nodule (solid component)-pleura attachment distance was significantly greater in tumors with VPI than in those without. Sun Q, et al[6]. named the multiple linear septations in pleural-attached nodules the 'jellyfish sign', which was confirmed in our study to potentially identify NSCLCs with VPI.

Table 3 | Comparisons of diagnostic capability values of each metric among six observers without (w/o) and with (w/) MIIF model's assistance

		AUC[a]				Accuracy[b]				Sensitivity[b]				Specificity[b]			
		internal	*P*	external	*P*	internal	*P*	external	*P*	internal	*P*	external	*P*	internal	*P*	external	*P*
	MIIF model	0.869		0.785		0.812		0.798		0.704		0.640		0.824		0.826	
Senior	Observer 1																
	w/o AI[c]	0.795	0.140	0.708	0.142	0.776	0.237	0.748	0.243	0.704	0.773	0.480	0.343	0.784	0.185	0.797	0.556
	w/ AI[d]	0.904	**0.012**	0.816	**0.013**	0.899	**<0.001**	0.840	**0.009**	0.778	0.724	0.720	**0.041**	0.912	**<0.001**	0.862	0.095
	Observer 2																
	w/o AI[c]	0.839	0.379	0.680	**0.035**	0.834	0.417	0.791	1.000	0.630	0.724	0.200	**0.003**	0.856	0.201	0.899	0.055
	w/ AI[d]	0.907	**0.030**	0.751	0.218	0.863	0.170	0.840	0.136	0.815	0.074	0.520	**0.013**	0.868	0.663	0.899	0.789
	Observer 3																
	w/o AI[c]	0.772	**0.021**	0.681	**0.034**	0.650	**<0.001**	0.546	**<0.001**	0.852	0.289	0.640	0.752	0.628	**<0.001**	0.529	**<0.001**
	w/ AI[d]	0.847	**0.037**	0.800	**0.008**	0.758	**<0.001**	0.742	**<0.001**	0.889	1.000	0.720	0.752	0.744	**<0.001**	0.746	**<0.001**
	Average																
	w/o AI[c]	0.870	0.972	0.713	0.111	0.798	0.671	0.755	0.296	0.815	0.505	0.480	0.343	0.796	0.349	0.804	0.677
	w/ AI[d]	0.919	0.105	0.822	**0.008**	0.866	**0.004**	0.847	**0.009**	0.889	0.683	0.680	0.182	0.864	**0.005**	0.877	**0.044**
Junior	Observer 4																
	w/o AI[c]	0.811	0.127	0.666	**0.027**	0.736	**0.007**	0.687	**0.007**	0.741	1.000	0.560	0.752	0.736	**0.002**	0.710	**0.006**
	w/ AI[d]	0.852	**0.006**	0.736	**0.007**	0.841	**<0.001**	0.791	**0.001**	0.815	0.480	0.600	1.000	0.844	**<0.001**	0.826	**<0.001**
	Observer 5																
	w/o AI[c]	0.782	**0.042**	0.724	0.226	0.856	0.067	0.810	0.855	0.704	0.724	0.400	0.077	0.872	**0.038**	0.884	0.136
	w/ AI[d]	0.873	**0.030**	0.825	**0.043**	0.884	0.099	0.877	**0.015**	0.778	0.683	0.640	**0.041**	0.896	0.149	0.920	0.228
	Observer 6																
	w/o AI[c]	0.767	**0.036**	0.656	**0.012**	0.451	**<0.001**	0.362	**<0.001**	0.889	0.182	0.880	**0.041**	0.404	**<0.001**	0.268	**<0.001**
	w/ AI[d]	0.863	**0.025**	0.770	**0.010**	0.617	**<0.001**	0.595	**<0.001**	0.963	0.480	0.920	1.000	0.580	**<0.001**	0.536	**<0.001**
	Average																
	w/o AI[c]	0.853	0.621	0.731	0.220	0.744	**0.010**	0.675	**0.003**	0.778	0.724	0.600	1.000	0.740	**0.002**	0.688	**0.002**
	w/ AI[d]	0.904	**0.036**	0.824	**<0.001**	0.838	**<0.001**	0.810	**<0.001**	0.852	0.617	0.680	0.683	0.836	**<0.001**	0.833	**<0.001**
Average																	
w/o AI[c]		0.879	0.765	0.739	0.266	0.736	**0.006**	0.663	**0.001**	0.889	0.131	0.600	1.000	0.720	**<0.001**	0.674	**<0.001**
w/ AI[d]		0.921	0.073	0.828	**0.003**	0.845	**<0.001**	0.828	**<0.001**	0.926	1.000	0.760	0.221	0.836	**<0.001**	0.841	**<0.001**

[a]Using the Delong tests.
[b]Using the McNemar tests.
[c]indicates that the diffetence between the diagnostic performance of the MIIF model and radiologists.
[d]indicates that the diffetence between the diagnostic performance of radiologists without and with MIIF model assistance.
Bold: statistically significant data.

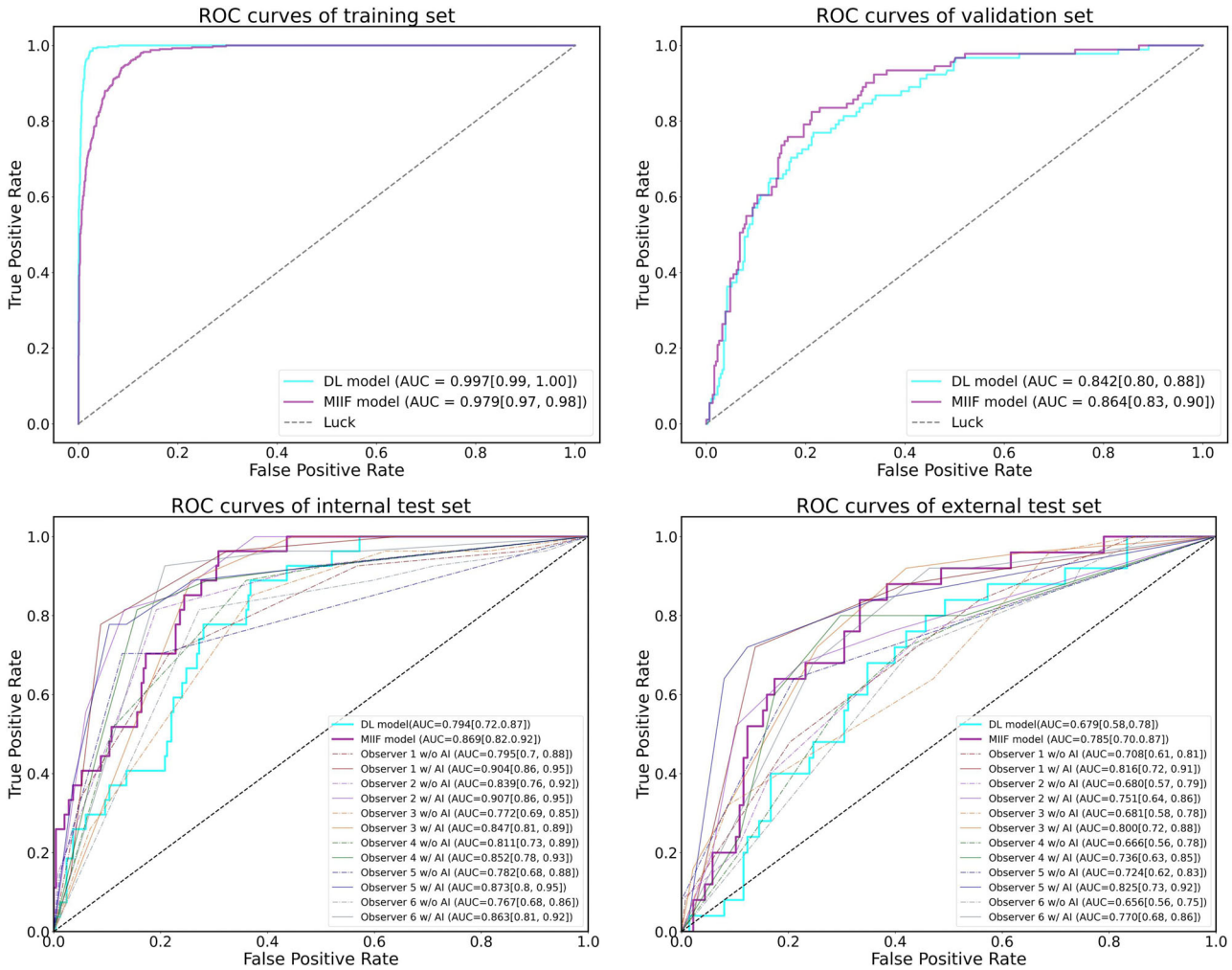

Fig. 3 | The receiver operating characteristic (ROC) curves of our proposed models in training, validation, internal and external test set.

In pleural-tag nodules, it was observed that those with VPI frequently exhibited pleural indentation (PI) (95.65% vs. 76.61%). Previously, the PIF value was defined to measure the degree of pleural indentation caused by the nodule[8]. In this larger study, the PIF value was identified as a significant independent predictor of VPI (OR, 23.32; $P = 0.003$). It is hypothesized that a higher PIF value may indicate a greater degree of intratumoral fibrosis, which could be associated with tumor invasiveness[7,9]. These specific CT findings can be utilized in daily clinical practice and have potential to predict VPI. Moreover, with the aid of our proposed MIIF model, the diagnostic performance improved.

Several limitations were noted in our study. First, the MIIF model demonstrated slightly inferior performance on the internal and external test sets compared to the training and validation sets. The lower incidence of VPI in clinical practice contributed to a small proportion of positive VPI cases in the training and validation sets, potentially affecting the robustness of our model. Second, the retrospective collection of data might lead to possible selection bias. Despite this, the datasets were compiled from a substantial patient cohort, encompassing internal and external test sets. Plans are in place to gather more data and integrate specific CT findings for further iteration and prospective validation of the MIIF model. Furthermore, cross-center variability in imaging protocols (e.g., reconstruction kernels and slice thickness) likely contributed to residual domain shift, as reflected by the lower external AUC. Although we standardized preprocessing (isotropic resampling and intensity normalization), we did not apply explicit domain-adaptation or harmonization in this study. Domain-adaptation/standardization methods-such as feature-space harmonization (e.g., ComBat) and image-level translation (e.g., CycleGAN)-are promising for reducing scanner- and protocol-related shifts[26,27]. In future work, we will systematically investigate and rigorously validate these methods, together with site-specific calibration, to improve robustness and narrow the performance gap in external cohorts.

In conclusion, the CT-based MIIF model developed for identifying VPI in NSCLC (with a solid component size of 30 mm or smaller) outperformed the diagnostic accuracy of radiologists, particularly for junior radiologists. Its high clinical utility could enhance their diagnostic efficacy. Specific CT findings, including SN, nodule solid component mean diameter, and PIF value, were identified as predictors of VPI and will be important features in the MIIF model.

Methods

The retrospective study received approval from the institutional review boards of all participating hospitals. The necessity for written informed consent was waived, as data were analyzed retrospectively and anonymously. Clinical and pathological data were reviewed from medical records, and CT images were obtained from picture archiving and communication system (PACS). Figure 5 illustrates a schematic drawing of the overall study design.

Study patients

The main dataset were collected from patients with small NSCLCs at six centers over earlier time windows for training and validating the proposed model: Center 1 (Shanghai Zhongshan Hospital, 2015.2–2018.11), Center 2

Table 4 | Comparisons of CT findings between nodules with and without VPI in the internal and external test sets

	Pleural-attached nodules				Pleural-tag nodules			
Characteristics	**VPI (24)[a]**	**non-VPI (136)**	**Total (160)**	***P* value**	**VPI (23)[a]**	**non-VPI (171)**	**Total (194)**	***P* value**
Nodule type				***<0.001***				**0.004**
PSN	4 (16.67%)	99 (72.79%)	103 (64.38%)		11 (47.83%)	130 (76.02%)	141 (72.68%)	
SN	20 (83.33%)	37 (27.21%)	57 (35.63%)		12 (52.17%)	41 (23.98%)	53 (27.32%)	
Mean diameter	19.75±5.08	15.69±5.85	16.30±5.91	***0.002***	19.53±4.84	15.58±4.79	16.05±4.94	**<0.001**
Nodule-pleura attachment distance (mm)	16.93±8.91	12.19±7.24	12.90±7.67	**0.014**				
Nodule-pleura distance (mm)					5.13±2.59	5.43±4.60	5.40±4.40	0.395
Indentation				0.893				**0.036**
positive	19 (79.17%)	106 (77.94%)	125 (78.13%)		22 (95.65%)	131 (76.61%)	153 (78.87%)	
negative	5 (20.83%)	30 (22.06%)	35 (21.88%)		1 (4.35%)	40 (23.39%)	41 (21.13%)	
Pleural tag type								**0.023**
I					0	42 (24.56%)	42 (21.65%)	
II					17 (73.91%)	102 (59.65%)	119 (61.34%)	0.581
III					6 (26.09%)	27 (15.79%)	33 (17.01%)	
PIF					0.51±0.13	0.29±0.22	0.32±0.22	**<0.001**
Jellyfish sign				**0.006**				
positive	11 (45.83%)	27 (19.85%)	38 (23.75%)					
negative	13 (54.17%)	109 (80.15%)	122 (76.25%)					
Bridge tag sign								0.403
positive					6 (26.09%)	32 (18.71%)	38 (19.59%)	
negative					17 (73.91%)	139 (81.29%)	156 (80.41%)	
Involved pleural type								
Interlobar pleura	1 (4.17%)	42 (30.88%)	43 (26.88%)		0	17 (9.94%)	17 (9.76%)	
non-Interlobar pleura	15 (62.50%)	69 (50.74%)	84 (52.50%)	0.434	19 (82.61%)	140 (81.87%)	159 (81.96%)	0.219
Complex pleura	8 (33.33%)	25 (18.38%)	33 (20.63%)		4 (17.39%)	14 (8.19%)	18 (9.28%)	
Solid component (PSN+SN)	23	103	126		21	128	149	
Mean diameter	19.37±5.60	12.24±6.86	13.54±7.18	**<0.001**	16.77±4.71	11.95±5.16	12.63±5.35	**<0.001**
Solid component-pleura attachment distance (mm)	16.65±9.10	9.73±7.14	11.00±7.96	**<0.001**				
Solid component-pleura distance (mm)					5.75±2.61	7.56±5.24	7.30±4.99	0.289

[a]Five nodules, which pushed against pleura, were all pathologically confirmed with VPI.
VPI visceral pleural invasion, *PSN* part solid nodule, *SN* solid nodule, *PI* pleural indentation, *PIF* pleural indentation fraction.
Bold: statistically significant data.

(Shanghai Public Health Clinical Center, 2015.6–2020.6), Center 3 (Shanghai Sixth People's Hospital, 2011.12–2019.10), Center 4 (Shanghai Ruijin Hospital, 2015.2–2021.8), Center 5 (Wuhan Union Hospital, 2014.9–2018.12) and Center 6 (Shanghai Xuhui District Central Hospital, 2017.8–2021.7).

A detailed flowchart of patient inclusion and exclusion was shown in Fig. 1. Patients were randomly assigned to the training and validation sets in an 8:2 ratio at the patient-level.

Consecutive patients underwent surgery for small NSCLC at Center 1 (Shanghai Zhongshan Hospital, 2018.12–2023.9) and Center 7 (Shanghai Minhang District Central Hospital, 2020.10–2024.2) were included as the internal and external test sets, respectively. The patient inclusion and exclusion criteria were consistent with those previously described (Fig. 1). Notably, the internal test temporal window at Center 1 does not overlap with Center 1's contribution to the main dataset, although it partially overlaps the overall multi-center timeframe of the main dataset.

Pathological analysis

All resected tumors were stained with Hematoxylin and Eosin, and their associated pleura were stained with masstone. Microscopic examination was conducted on specimens sliced at 0.5 cm thickness. Elastica van Gieson staining was applied whenever the elastic layer of the involved visceral pleura was indistinct. VPI was defined as tumor invasion beyond the elastic layer, classified as Pl1 (without exposure on the pleural surface) or Pl2 (with exposure), but excluding involvement of the parietal pleura[28].

Data preparation

Detailed CT acquisition parameters for each participating center were provided in Supplementary Table 4. Tumor volumes of interest from all datasets were automatically segmented on CT images using the research platform uAI Research Portal (uRP, United Imaging Intelligence Co., Ltd.)[29]. A three-dimensional (3D) DL network, VB-net, was utilized to automatically detect and segment lung tumors. This automated approach achieved a Dice similarity coefficient of 91.5%[30]. Radiologist S.Y.Y. subsequently reviewed and manually adjusted the segmentation results as needed, utilizing the lung window setting (window center = −450 to -600 Hu, window width = 1500 to 2000 Hu) on uRP.

Model development

A MIIF model for VPI prediction was proposed in this study. The MIIF model comprises three principal components: 1) Image preprocessing, including imaging cropping, resampling, and normalization; 2) Multi-feature extraction, where an attention-based residual network was trained on the training set to extract 256 deep image features from the last fully

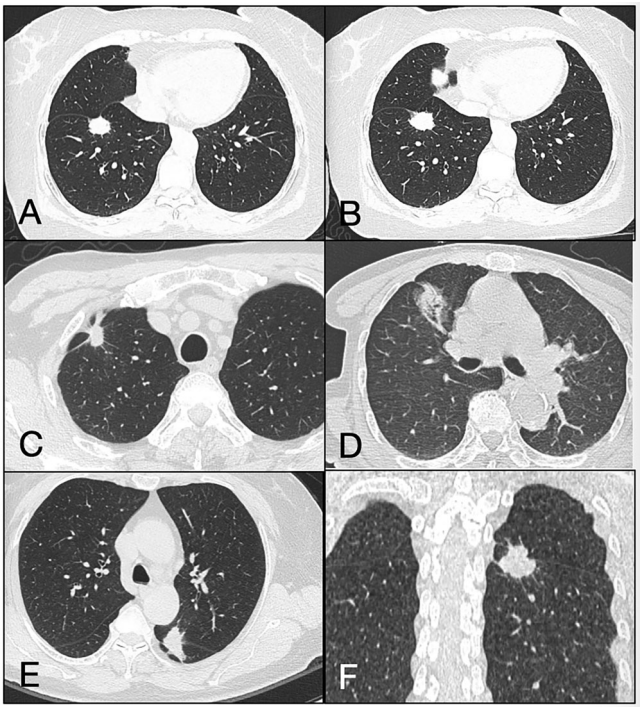

Fig. 4 | Schematic illustration of pleura-associated nodules on CT. Representative images of the pleural-attached nodule (**A**, **B**), pleural-tag nodule (**C**, **D**) and nodule pushed against pleura (**E**, **F**). Different slice of the same nodule in CT image of a right lower lobe pleural-attached nodule with pleural indentation (**A**, **B**). The nodule surface directly touched the pleura surface. Axial CT (**C**, **D**) showed a right upper lobe pleural-tag nodule. There was one or more linear tag between the nodule and pleura, but the nodule does not directly touch the pleura. Axial CT (**E**) and coronal CT (**F**) showed a left lower lobe nodule pushed against the interlobar pleura.

connected layer. Additionally, 1185 radiomics features were extracted using PyRadiomics (version 3.0.1), and 13 general CT findings were automatically obtained and verified; and 3) MIIF model construction, where the deep imaging features, radiomics, and general CT findings were integrated. Z-score normalization, the LASSO algorithm, and a machine learning classifier (quadratic discriminant analysis, QDA) were employed for model construction.

The enrolled CT volumes were first resampled into $0.7 \times 0.7 \times 1.0\ mm^3$ resolution by trilinear interpolation, and cropped around the centers of lung nodules with a cubic patch of $112 \times 112 \times 96$ voxels. Then the CT intensity was converted into Hounsfield units (HU) and normalized by Z-score standardization method. Finally, the image intensity of each patient was clipped to the range of $[-1, 1]$ to facilitate the observation of lung tissue. The equation of CT normalization is given as follows:

$$I = \begin{cases} -1, if\ \frac{I_{Hu} - mean}{STD} < -1 \\ 1, if\ \frac{I_{Hu} - mean}{STD} > 1 \\ \frac{I_{Hu} - mean}{STD}, other \end{cases} \quad (1)$$

Where the mean value is set as -400 and the STD (standard deviation) is set as 750. This process ensures that each CT scan is standardized to have a uniform resolution and a range of the same intensity.

We developed a DL model to predict VPI status and extract deep imaging features. The DL model incorporates residual blocks and a class activation mapping (CAM) mechanism, enabling it to focus on the nodule and its surrounding pleural region. The architecture of the DL model consists of two major components: image augmentation; and an attention-based nodule diagnosis module with CAM. The model with the best performance on the validation set was selected, and 256 deep imaging features were extracted from its last fully connected layer.

In this study, we randomly adopted flipping along each axis, scaling by a range of 0.8 to 1.2, and rotation by an angle along an axis in a range of $-10°$ to $10°$ on each CT patches in the training set with a probability of 50%. Besides, the number of VPI-negative and VPI-positive data in the training samples is extremely imbalanced, with a distribution ratio close to 4:1. The imbalance in samples can lead to significant bias in our classification model. Therefore, in this study, we address this issue by conducting over-sampling of the minority class to increase the number of samples in the minority class, aiming to achieve a balanced ratio of positive and negative samples in the input network.

The DL model was built on the foundation of 3D residual network (ResNet) framework, serving to discriminate VPI-positive from VPI-negative in NSCLC. To guide the network to focus on the features of NSCLC and its surrounding regions, CAM attention mechanism was introduced in this framework. Online supervision of network response regions was implemented during the training procedure to optimize the classification performance.

The architecture consists of several key components, including an input block, four downsampling blocks, a global average pooling (GAP) layer, a fully connected layer and a softmax layer. Specifically, the input block consists of a convolutional module that includes a $3 \times 3 \times 3$ kernel size and a $1 \times 1 \times 1$ stride size convolutional layer, followed by a batch normalization (BN) layer and a ReLU layer. For the downsampling blocks, we employed residual structures where the input and output of each downsampling block are combined through addition and then passed as input to the next downsampling block. Each downsampling block begins with a convolutional block that includes a convolutional layer (kernel size=$2 \times 2 \times 2$, stride size= $2 \times 2 \times 2$), followed by a BN layer and a ReLU layer. The remaining convolutional blocks in each downsampling block consist of a convolutional layer with kernel size of $3 \times 3 \times 3$ and stride size of $1 \times 1 \times 1$, along with subsequent a BN and ReLU layer. The output channels for the input block and the four downsampling blocks are set as follows: 16, 32, 64, 128 and 256, respectively. The softmax layer has an output channel setting of 2 which represents the probabilities of VPI-negative and VPI-positive for the given input sample.

In order to enable the model to learn feature related to nodules and their pleural region associated with VPI, this study utilizes attention maps of 3D CAM for online learning. Specifically, a $1 \times 1 \times 1$ 3D convolutional layer and ReLU activation function were used to generate attention feature maps corresponding to network response regions. The equation of attention maps are as follows:

$$A = ReLU(Convolution(f, w)) \quad (2)$$

Where f represents the feature map before the GAP layer, w represents the weight matrix of the fully connected layer. To make our attention generation procedure trainable, a convolution layer with kernel size of 1×1×1 and a ReLU layer was employed to generate the attention feature map of A. The size of A is 1/16 of the corresponding size of the input CT images. We then upsampled the attention feature map to match the input image size, normalized it to a range of 0-1, and applied a sigmoid function for linear mapping to obtain the final attention map, as follows:

$$T(A) = Sigmoid(A) = \frac{1}{1 + exp(-\alpha(A - \beta))} \quad (3)$$

The values of α and β are set to 100 and 0.4, respectively, where $T(A)$ represents the attention map generated by the online attention module. During the training process, the weights of the fully connected layer are used to assign values to the convolution kernel parameters in the Eq.2, thereby optimizing the online network attention map.

The loss function of the proposed DL model includes label smooth cross entropy (LSCE) loss and mean squared error (MSE) loss, which can be

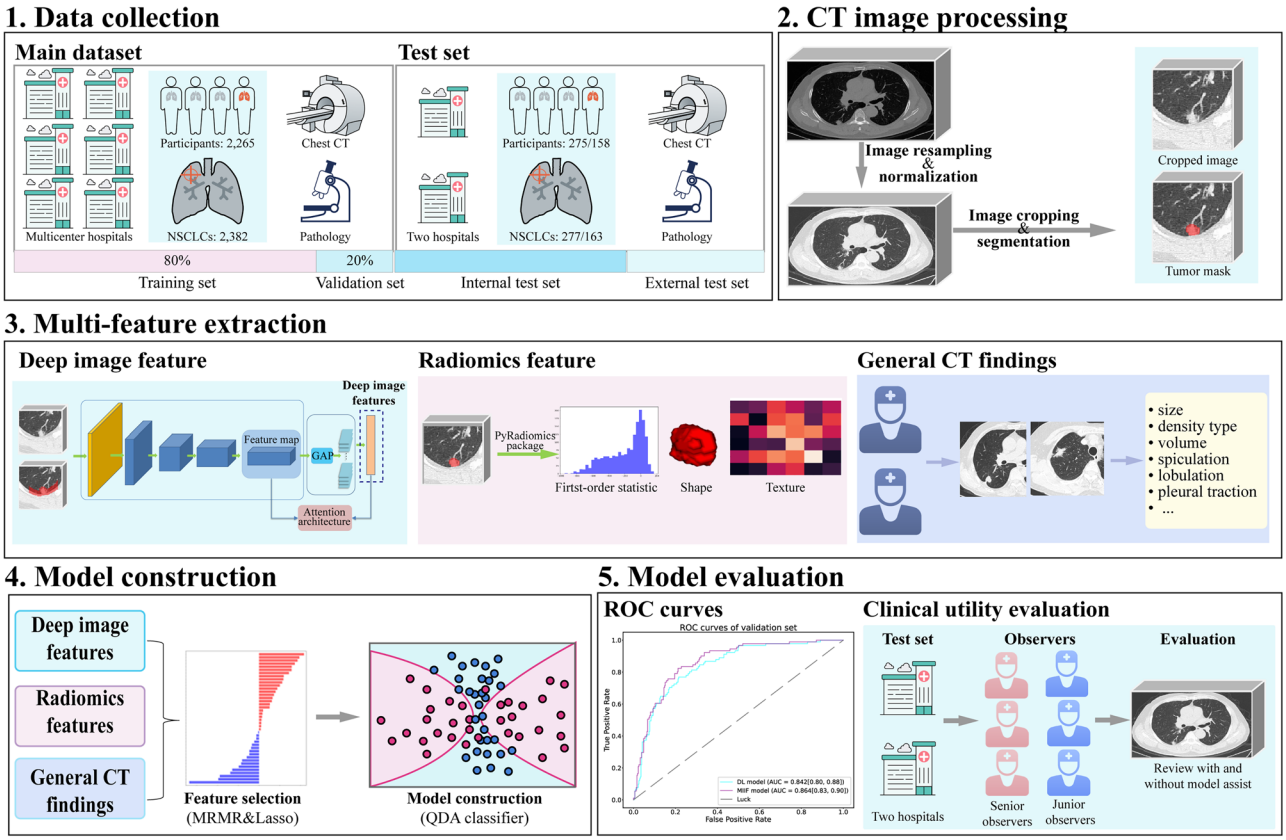

Fig. 5 | Schematic drawing of the overall study design.

rewritten as follows:

$$\text{Loss} = \text{LSCE\_loss} + \alpha * \text{MSE\_loss} \tag{4}$$

$$\text{LSCE\_loss} = -\sum_{i=1}^{N} q_i \log y_i \tag{5}$$

$$q_i = \begin{cases} 1-\varepsilon if\, i == y, \\ \varepsilon/(\text{N}-1) otherwise, \end{cases} \tag{6}$$

$$\text{MSE\_loss} = \frac{1}{n}\sum_{i=1}^{n}(T(A)-G)^2 \tag{7}$$

Where q_i denotes the positive prediction probability of the input samples, y_i is the gold standard, N is the number of samples, and ε is a small constant, which makes the target probability in the LSCE loss function no longer 1 or 0, thus avoiding overfitting of the model. $T(A)$ represents the network attention map of the network, and G is the input volume of interest (VOI). The MSE loss function is used to make the attention map of the network as similar as possible to the input VOI, thereby achieving guidance of the attention regions in the network. In the training process, a constant weight α is used to balance the classification task and the task of guiding the attentional map of the network.

We trained the attention-based 3D ResNet for a maximum of 1001 epochs with early stopping (patience of 100 epochs). All network parameters are initialized by Kaiming initialization. The model was optimized using the Adam optimizer with an initial learning rate of 10^{-5}, betas of 0.9 to 0.999, epsilon of 1×10^{-8}, and weight decay of 0.01. The learning rate was scheduled using MultiStepLR with milestones at epochs 50, 100, 200, 400, and 800, and a gamma value of 0.4. The batch size was set to 36. A total of 16 CPU threads were used for data loading. The smoothing factor (ε) of LSCE loss is set as 0.1, and the attention guidance weight (α) of MSE loss is set as 100. No dropout was applied in the network architecture. The model with the best performance on the validation set was finally selected.

A total of 1185 radiomics features were extracted from the preprocessed CT images using the PyRadiomics (version 3.0.1) package implemented in the research platform uAI Research Portal (uRP, United Imaging Intelligence Co., Ltd.). Three types of features fell into categories: first-order statistics, morphological features, and texture features. They were extracted from 3D tumor masks in the original CT images, and from two filtered images (i.e., Laplacian of Gaussian and wavelet filtered images).

Thirteen general CT findings were automatically obtained for each pulmonary nodule through the uRP platform, including density, size, volume, long and short diameters, and nodule signs of spiculation, vocule sign, calcification, lobulation, pleural traction (traditionally, pleural tag sign/pleural indentation), air bronchogram, spinous, and vessel convergence. All CT findings were manually verified by a radiologist (S.Y.Y) and corrected if necessary.

We integrated the deep imaging features, radiomics features, and general CT findings, resulting in 1454 features per pulmonary nodule. To ensure comparability across different feature types, Z-score normalization were applied. Subsequently, the LASSO algorithm was employed to select the most relevant features for distinguishing between positive VPI and negative VPI nodules. The LASSO regularization was performed with L1 penalty using 5-fold stratified cross-validation. We conducted a grid search over the regularization strength α in the range [0.001, 0.01, 0.05, 0.1, 1, 10]. Features that were selected in at least 80% of the cross-validation folds were retained. The selected features (n = 42) were then used to train a QDA classifier. The QDA classifier was implemented with a regularization parameter of 0.1 to prevent overfitting by smoothing the covariance estimates. The priors were set based on the class frequencies in the training set.

All experiments were conducted on a high-performance computing cluster with NVIDIA A40 GPUs (48GB memory). We used Pytorch1.12.1,

CUDA 12.4, and scikit-learn 1.6.1. Random seeds were fixed for PyTorch, NumPy, and Python to ensure deterministic results.

To mitigate class imbalance, several strategies were applied to handlie class imbalance: 1) Oversampling of the minority class. During DL model training, we oversampled VPI-positive cases to achieve a 1:1 ratio per batch, ensuring the network learned balanced features; 2) Loss function design. We used LSCE, which down-weights overconfident predictions and improves minority-class learning; 3) Stratified cross-validation during LASSO. We used LASSO for feature selection on the general CT/radiomics/deep learning features prior to model fusion. Stratified 5-fold cross-validation was performed, i.e., each fold preserved the proportion of VPI-positive and VPI-negative cases present in the full training set. Stratification reduces variability in class composition across folds under imbalance, yielding a more stable selection of regulation parameter of LASSO and preserving minority-informative features; 4) QDA priors set to training frequencies. The MIIF classifier uses QDA. QDA models class-conditional densities with class-specific covariance matrices and uses Bayes' rule. Aligning priors with the development base rate mitigates bias from assuming equal priors in an imbalanced setting and improves probability calibration within development-like distributions.

Observer performance test

Six board-certified thoracic radiologists (Y.Z., S.Y.Y., Q.W., W.S., S.Y., F.S., with 6-23 years of experience in chest imaging) independently assessed the presence of VPI in NSCLC using a 5-point scale: 0-unlikely to have VPI (0% possibility); 1-slight likely to have VPI (0-25% possibility); 2-moderately likely to have VPI (26-50% possibility); 3-very likely to have VPI (51-75% possibility); 4-extremely likely to have VPI (76-100% possibility). A CT-VPI presence score of 3 or 4 defined the presence of VPI. The radiologists were informed of the patients' ID, age, and tumor location. To mirror daily clinical practice, scoring was conducted without prior education on specific pleural-related CT findings. Thus, the CT-VPI presence score was determined based on the radiologists' own experience. The six radiologists assessed the CT-VPI presence using axial, coronal, and sagittal images for all patients in the internal and external test sets to evaluate the relationship between the tumors and pleura.

Paired design (sequential session) for comparing diagnostic performance

To compare performance between AI-unassisted and AI-assisted interpretations, interpretation typically occurs without AI in the first session and with AI in the second session[11]. In this study, a washout period (> one month) was implemented between the two sessions to prevent learning effects from the first session. The order of case review was randomly reshuffled prior to the second session. A 5-point scale was also employed to evaluate the likelihood of VPI presence on CT with the MIIF model results, conducted by the same 6 radiologists.

Specific CT findings evaluation

For the internal and external test sets, tumors were categorized into SN, PSN, and GGN based on CT image analyses. Subpleural nodules identified by CT were classified into three categories (Fig. 4):

1. Pleural-attached nodules, which were in direct contact with the pleural surface.
2. Pleural-tag nodules, which were not in contact with the pleura[6]. The nodules were with thin, linear structures (≤2 mm in maximum width) extending from the surface of nodule to the visceral pleura; the tag must be continuous with both the nodule and the pleura (to distinguish it from unrelated linear opacities such as atelectatic bands).
3. Nodules that pushed against the pleura, which were also in direct contact with the interlobar fissure. The pulmonary nodule displaced the pleura to the opposite side, or growed across the fissure.

For pleural-attached nodule, specific CT findings were assessed, including the distance between the nodule (solid component) and the pleura, PI, and the jellyfish sign. For pleural-tag nodules, specific CT findings such as the distance between the nodule (solid component) and the pleura, bridge tag sign, PI, PIF, and pleural tag type (Supplementary Fig. 2) were also evaluated.

The CT findings for each subpleural nodule were assessed by two radiologists (S.Y.Y./F.S., with 10/23 years of experience in chest CT imaging), who were blinded to the clinicopathologic data. Any disagreements were evaluated together and resolved by consensus.

Statistical analysis

Model performance was evaluated using AUC, accuracy, sensitivity, specificity, positive and negative predictive values (PPV, NPV), and F1-score. Calibration curves and DCA were also performed to evaluate the accuracy of risk estimate. Additionally, Brier scores were calculated that quantitatively measure the distance in the probability domain and a lower score means better prediction. For both the internal and external test sets, AUC comparisons between the MIIF model and radiologists, and between unassisted and assisted radiologist interpretations, were performed using the DeLong test. AUC interpretations were as follows: 1) acceptable (AUC, 0.70–0.80), 2) excellent (AUC, 0.80–0.90), 3) outstanding (AUC, greater than 0.90)[31]. Continuous variables were presented as mean ± standard deviation and analyzed using the independent samples t-test, Mann-Whitney *U* test, or analysis of variance, depending on data distribution. Categorical variables were presented as frequencies with percentages and analyzed using the Pearson χ^2 or Fisher exact test, as appropriate. Multivariable logistic regression analysis (Forward: LR) was utilized to identify independent CT features associated with VPI. The McNemar test was employed to compare parameters of accuracy, sensitivity, and specificity. Analyses were conducted using SPSS software (version 29.0; IBM) and Python (version 3.9.12). A two-tailed *P*-value of less than 0.05 was considered statistical significant.

Sample size calculation

We calculated the required sample size for developing the prediction model for VPI status using the approach described by Riley et al.[32]. This method aims to minimize model overfitting and ensure precise predictions by considering multiple criteria. The sample size was determined using the following formula for binary outcomes:

$$n = \left(\frac{Z}{\sigma}\right)^2 \hat{\Phi}(1-\hat{\Phi})$$

where n represents required sample size, Z refers to Z-value,σ refers to the margin of error and is generally recommended as ≤0.05, $\hat{\Phi}$ is the overall outcome proportion. In this study, the Z is set as 1.96 for a 95% confidence level, σ is set as 0.05. Based on previous study by Huang et al.[33], the VPI incidence rate for lung tumors less than 30 mm in size ranges from 8% to 38%. We calculated the sample size for both the lower and upper bounds of this range: at least 114 participants (i.e., about 10 participants with positive VPI) are required when the outcome proportion ($\hat{\Phi}$) is 0.08, and 363 participants (i.e., 138 participants with positive VPI) are required when the $\hat{\Phi}$ is 0.38.

To ensure robust model development and validation, we aimed to at least recruit a sample size of 363 participants (including approximately 138 with positive VPI) at the upper end of the incidence rate range.

Data availability

No datasets were generated or analysed during the current study.

Received: 31 March 2025; Accepted: 20 January 2026;
Published online: 29 January 2026

References

1. Lung cancer [https://www.who.int/news-room/fact-sheets/detail/lung-cancer].

2. Rami-Porta, R. et al. The International Association for the Study of Lung Cancer Lung Cancer Staging Project: proposals for revision of the TNM stage groups in the forthcoming (Ninth) edition of the TNM classification for lung cancer. *J. Thorac. Oncol.* **19**, 1007–1027 (2024).
3. Gorai, A. et al. The clinicopathological features associated with skip N2 metastases in patients with clinical stage IA non-small-cell lung cancer. *Eur. J. Cardiothorac. Surg.* **47**, 653–658 (2015).
4. Kudo, Y. et al. Impact of visceral pleural invasion on the survival of patients with non-small cell lung cancer. *Lung Cancer* **78**, 153–160 (2012).
5. Altorki, N. et al. Recurrence of non-small cell lung cancer with visceral pleural invasion: a secondary analysis of a randomized clinical trial. *JAMA Oncol.* **10**, 1179–1186 (2024).
6. Sun, Q. et al. CT predictors of visceral pleural invasion in patients with non-small cell lung cancers 30 mm or smaller. *Radiology* **310**, e231611 (2024).
7. Onoda, H. et al. Correlation between pleural tags on CT and visceral pleural invasion of peripheral lung cancer that does not appear touching the pleural surface. *Eur. Radio.* **31**, 9022–9029 (2021).
8. Yang, S. et al. Visceral pleural invasion by pulmonary adenocarcinoma ≤3 cm: the pathological correlation with pleural signs on computed tomography. *J. Thorac. Dis.* **10**, 3992–3999 (2018).
9. Hsu, J. S. et al. Pleural tags on CT scans to predict visceral pleural invasion of non-small cell lung cancer that does not abut the pleura. *Radiology* **279**, 590–596 (2016).
10. Li, M. et al. Preoperative prediction of peritoneal metastasis in colorectal cancer using a clinical-radiomics model. *Eur. J. Radio.* **132**, 109326 (2020).
11. Park, S. H. et al. Methods for clinical evaluation of artificial intelligence algorithms for medical diagnosis. *Radiology* **306**, 20–31 (2023).
12. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. & Aerts, H. Artificial intelligence in radiology. *Nat. Rev. Cancer* **18**, 500–510 (2018).
13. Lin, X. et al. A CT-based deep learning model: visceral pleural invasion and survival prediction in clinical stage IA lung adenocarcinoma. *iScience* **27**, 108712 (2024).
14. Choi, H. et al. Prediction of visceral pleural invasion in lung cancer on CT: deep learning model achieves a radiologist-level performance with adaptive sensitivity and specificity to clinical needs. *Eur. Radio.* **31**, 2866–2876 (2021).
15. Lim, W. H. et al. Diagnostic performance and prognostic value of CT-defined visceral pleural invasion in early-stage lung adenocarcinomas. *Eur. Radio.* **34**, 1934–1945 (2024).
16. Jin, J. et al. Deep learning radiomics model accurately predicts hepatocellular carcinoma occurrence in chronic hepatitis B patients: a five-year follow-up. *Am. J. Cancer Res*. **11**, 576–589 (2021).
17. Braghetto, A., Marturano, F., Paiusco, M., Baiesi, M. & Bettinelli, A. Radiomics and deep learning methods for the prediction of 2-year overall survival in LUNG1 dataset. *Sci. Rep.* **12**, 14132 (2022).
18. Wo, Y. et al. Impact of visceral pleural invasion on the association of extent of lymphadenectomy and survival in stage I non-small cell lung cancer. *Cancer Med.* **8**, 669–678 (2019).
19. Heidinger, B. H. et al. Visceral pleural invasion in pulmonary adenocarcinoma: differences in CT patterns between solid and subsolid cancers. *Radio. Cardiothorac. Imaging* **1**, e190071 (2019).
20. Elicker, B. M. Pleural invasion in subsolid and solid lung cancers: predictive features at CT and their clinical significance. *Radio. Cardiothorac. Imaging* **1**, e190145 (2019).
21. Shi, J. et al. The combination of computed tomography features and circulating tumor cells increases the surgical prediction of visceral pleural invasion in clinical T1N0M0 lung adenocarcinoma. *Transl. Lung Cancer Res.* **10**, 4266–4280 (2021).
22. Zhao, L. L. et al. Visceral pleural invasion in lung adenocarcinoma ≤3 cm with ground-glass opacity: a clinical, pathological and radiological study. *J. Thorac. Dis.* **8**, 1788–1797 (2016).
23. Bankier, A. A. et al. Recommendations for measuring pulmonary nodules at CT: a statement from the Fleischner Society. *Radiology* **285**, 584–600 (2017).
24. Riely, G. J. et al. Non-Small Cell Lung Cancer, Version 4.2024, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc. Netw.* **22**, 249–274 (2024).
25. Yang, Y. et al. Using CT imaging features to predict visceral pleural invasion of non-small-cell lung cancer. *Clin. Radio.* **78**, e909–e917 (2023).
26. Fortin, J. P. et al. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* **161**, 149–170 (2017).
27. Zhu J. Y., Park T., Isola P., Efros A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proc. IEEE Int. Conf. Comput. Vis.* 2242-2251 (2017).
28. Detterbeck, F. C., Boffa, D. J., Kim, A. W. & Tanoue, L. T. The Eighth Edition Lung Cancer Stage Classification. *Chest* **151**, 193–203 (2017).
29. Wu, J. et al. uRP: An integrated research platform for one-stop analysis of medical images. *Front Radio.* **3**, 1153784 (2023).
30. Chen, L. et al. An artificial-intelligence lung imaging analysis system (ALIAS) for population-based nodule computing in CT scans. *Comput. Med. Imaging Graph*. **89**, 101899 (2021).
31. Mandrekar, J. N. Receiver operating characteristic curve in diagnostic test assessment. *J. Thorac. Oncol.* **5**, 1315–1316 (2010).
32. Riley, R. D. et al. Calculating the sample size required for developing a clinical prediction model. *Bmj* **368**, m441 (2020).
33. Huang, H., Wang, T., Hu, B. & Pan, C. Visceral pleural invasion remains a size-independent prognostic factor in stage I non-small cell lung cancer. *Ann. Thorac. Surg.* **99**, 1130–1139 (2015).

Acknowledgements

We would like to acknowledge Yuehua Li (Shanghai Sixth People's Hospital), Zenghui Cheng (Shanghai Ruijin Hospital), Heshui Shi (Wuhan Union Hospital), Yonghua Xu (Shanghai Xuhui District Central Hospital) and Bin Song (Shanghai Minhang District Central Hospital), for clinical and imaging data collection. This work was supported by the National Natural Science Foundation of China (82172030), Science and Technology Commission of Shanghai Municipality (22YF1443500), Shanghai Municipal Hospital Development Center (SHDC2020CR3080B).

Author contributions

S.Y.Y., Z.Z. and F.S. conceived and designed the study. S.Y.Y., Q.W., Y.Y.Z., S.Y., W.S., Y.G., T.C., Y.Z., H.Z., F.S., Y.S., M.Z. collected the data. Y.W., Y.H. and F.S. developed and validated the MIIF model. S.Y.Y. and Y.W. wrote the original manuscript. Z.X., Z.Z. and F.S. reviewed and edited the manuscript. All authors approved the final version of the manuscript and had the final responsibility for the decision to submit for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-026-01305-4.

Correspondence and requests for materials should be addressed to Zhong Xue, Zhiyong Zhang or Fei Shan.