Article

Whole slide image based deep learning refines prognosis and therapeutic response evaluation in lung adenocarcinoma

Check for updates

Tao Chen[1,9], Jialiang Wen[1,9], Xinchen Shen[1,9], Jiaqi Shen[2], Jiajun Deng[1], Mengmeng Zhao[1], Long Xu[1], Chunyan Wu[3], Bentong Yu[4], Minglei Yang[5], Minjie Ma[6], Junqi Wu[1], Yunlang She [1], Yifan Zhong[1] ✉, Likun Hou[3] ✉, Yanrui Jin [7,8] ✉ & Chang Chen [1] ✉

Existing prognostic models are useful for estimating the prognosis of lung adenocarcinoma patients, but there remains room for improvement. In the current study, we developed a deep learning model based on histopathological images to predict the recurrence risk of lung adenocarcinoma patients. The efficiency of the model was then evaluated in independent multicenter cohorts. The model defined high- and low-risk groups successfully stratified prognosis of the entire cohort. Moreover, multivariable Cox analysis identified the model defined risk groups as an independent predictor for disease-free survival. Importantly, combining TNM stage with the established model helped to distinguish subgroups of patients with high-risk stage II and stage III disease who are highly likely to benefit from adjuvant chemotherapy. Overall, our study highlights the significant value of the constructed model to serve as a complementary biomarker for survival stratification and adjuvant therapy selection for lung adenocarcinoma patients after resection.

Estimating prognosis is essential for adjuvant treatment decision making and follow-up strategy selection for lung adenocarcinoma patients after surgery[1]. Some pathological factors, such as visceral pleural invasion (VPI)[2], spread through the air space (STAS)[3], and lymphovascular invasion (LVI)[4], have been reported to be associated with patient outcomes. Compared to the factors mentioned above, the International Association for the Study of Lung Cancer (IASLC)-proposed grading system has been proven to be more efficient and robust for patient stratification according to refs. 5,6. However, these factors may affect the prognosis of patients with stage I tumors, but their effect on those with stage II or III tumors requires further investigation. Furthermore, the TNM staging system can be used to categorize patients into several groups with distinct survival outcomes[7]. Nevertheless, there is often variation in patient outcomes

even among those at a specific TNM stage. Above all, there remains room for improvement in precise risk stratification to improve patient management and disease outcomes.

Recent advances in artificial intelligence (AI) have enabled the use of quantitative data derived from whole slide images (WSIs) to predict patient outcomes directly[8–10]. Histopathology images contain prognostically important information such as tumor-infiltrating lymphocytes[11,12], and proportions of tissue types[13], each of which can be quantified by specific digital pathology approaches. The hidden information in routine haematoxylin and eosin (H&E)-stained images may help to stratify prognosis from a different dimension, and may serve as a complementary biomarker to the current clinical variables.

In the present study, we developed a WSI-based deep learning model to predict the recurrence risk of resected lung adenocarcinoma without any

[1]Department of Thoracic Surgery, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai, China. [2]School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, China. [3]Department of Pathology, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai, China. [4]Department of Thoracic Surgery, The First Affiliated Hospital of Nanchang University, Jiangxi, China. [5]Department of Thoracic Surgery, Ningbo HwaMei Hospital, Chinese Academy of Science, Zhejiang, China. [6]Department of Thoracic Surgery, The First Affiliated Hospital of Lanzhou University, Gansu, China. [7]State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China. [8]MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai, China. [9]These authors contributed equally: Tao Chen, Jialiang Wen, Xinchen Shen. ✉e-mail: yifanzhong@126.com; hlk9575@163.com; jinyanrui@sjtu.edu.cn; chenthoracic@163.com

annotations from pathologists. We then assessed the ability of our model to stratify patients according to prognosis and investigated whether it could help refine the populations of patients likely to benefit from adjuvant chemotherapy. Finally, we obtained WSI heatmaps to explore the pathological features that may contribute to the predictive value of the model and the underlying biological basis of the model was also explored.

Results

Patient characteristics
With the constructed model, a WSI-based score was calculated for each patient in the two validation sets (Fig. 1). Patients in both validation sets were divided into low- and high-risk groups according to the median WSI-based score. In validation cohort 1, more patients in the high-risk group had STAS (36.3% vs. 29.3%, $p = 0.048$), IASLC grade III tumors (61.5% vs. 45.5%, $p < 0.001$), and adjuvant chemotherapy (39.7% vs. 30.4%, $p = 0.010$). In addition, more patients in the high-risk group had VPI (26.6% vs. 21.3%, $p = 0.098$), LVI (15.9% vs. 12.2%, $p = 0.163$), and TNM stage III tumors (17.3% vs. 12.8%, $p = 0.234$), although these differences were not statistically significant (Table 1). In validation cohort 2, the proportions of patients with VPI (35.6% vs. 23.9%, $p = 0.089$), STAS (39.1% vs. 34.1%, $p = 0.493$), LVI (21.8% vs. 12.5%, $p = 0.101$), IASLC grade III tumors (57.5% vs. 53.4%, $p = 0.826$), TNM stage III tumors (20.7% vs. 14.8%, $p = 0.382$), and adjuvant chemotherapy (34.5% vs. 29.5%, $p = 0.484$) were higher in the high-risk group, but none of the differences were statistically significant (Table 1).

Survival analysis of DFS
In validation cohort 1, the model successfully stratified patients into high- and low-risk groups according to prognosis in the entire cohort (hazard ratio [HR] 1.95, 95% confidence interval [CI] 1.46–2.62, $p < 0.001$) and in most of the prespecified subgroups (Fig. 2a). Similar results were found in the analysis of validation cohort 2 (Fig. 2b).

The incremental prognostic value of the WSI-based score
Univariable analysis of the validation cohort 1 revealed that patient outcomes were strongly associated with pathological-related factors, including VPI ($p < 0.001$), STAS ($p < 0.001$), and LVI status ($p < 0.001$), IASLC grade ($p < 0.001$), TNM stage ($p < 0.001$), and WSI-based score ($p < 0.001$) (Table 2). Before incorporating the WSI-based score into the multivariable Cox model, the IASLC grade (grade II vs. grade I, HR 13.17, 95% CI 3.16–54.96, $p < 0.001$; grade III vs. grade I, HR 26.99, 95% CI 6.50–112.13, $p < 0.001$) and TNM stage (stage II vs. stage I, HR 1.59, 95% CI 0.98–2.57, $p = 0.061$; stage III vs. stage I, HR 4.04, 95% CI 2.83–5.76, $p < 0.001$) were identified as independent predictors of DFS. After incorporating the WSI-based score into the multivariable Cox model, it is suggested that the IASLC grade (grade II vs. grade I, HR 13.54, 95% CI 3.24–56.52, $p < 0.001$; grade III vs. grade I, HR 25.69, 95% CI 6.19–106.63, $p < 0.001$), TNM stage (stage II vs. stage I, HR 1.67, 95% CI 1.03–2.71, $p = 0.037$; stage III vs. stage I, HR 4.22, 95% CI 2.95–6.05, $p < 0.001$), and the constructed model (HR, 1.82, 95% CI, 1.35–2.44, $p < 0.001$) were all independent predictors of DFS (Table 2). Similar results were found in the analysis of validation cohort 2 (Table 3).

We then used the C-index to compare the performance of each variable for predicting DFS. For variables significantly associated with DFS according to univariable Cox regression analysis, the WSI-based score did not outperform several pathological factors in either validation cohort 1 (C-index $_{[WSI-based\ score]}$ = 0.586; C-index $_{[IASLC\ grade]}$ = 0.674; C-index $_{[TNM\ stage]}$ = 0.665) or validation cohort 2 (C-index $_{[WSI-based\ score]}$ = 0.643; C-index $_{[VPI]}$ = 0.654; C-index $_{[STAS]}$ = 0.651; C-index $_{[IASLC\ grade]}$ = 0.718; C-index $_{[TNM\ stage]}$ = 0.699) (Supplementary Table 1; Supplementary Fig. 1). Regarding the 3-year and 5-year AUCs for predicting DFS, our model did not show an advantage over some other pathological factors (Supplementary Table 1; Supplementary Fig. 1).

For variables independently predicting DFS in the multivariable Cox regression analysis, we compared the predictive performance of their combinations. The results showed that the combination of IASLC grade, TNM stage

and WSI-based score (C-index $_{[WSI-based\ score\ \&\ IASLC\ grade\ \&\ TNM\ stage]}$ = 0.753) outperformed any combination of two variables in validation cohort 1 (C-index $_{[IASLC\ grade\ \&\ TNM\ stage]}$ = 0.737, $p < 0.001$; C-index $_{[WSI-based\ score\ \&\ IASLC\ grade]}$ = 0.708, $p < 0.001$; C-index $_{[WSI-based\ score\ \&\ TNM\ stage]}$ = 0.706, p < 0.001; Fig. 3a; Table 4). In validation cohort 2, the combination of three variables (C-index $_{[WSI-based\ score\ \&\ IASLC\ grade\ \&\ TNM\ stage]}$ = 0.811) also outperformed any combination of two variables (C-index $_{[IASLC\ grade\ \&\ TNM\ stage]}$ = 0.777, $p < 0.001$; C-index $_{[WSI-based\ score\ \&\ IASLC\ grade]}$ = 0.786, $p < 0.001$; C-index $_{[WSI-based\ score\ \&\ TNM\ stage]}$ = 0.763, $p < 0.001$; Fig. 3d; Table 4). The combined model also showed advantages with respect to the 3-year and 5-year AUCs for predicting DFS (Fig. 3b, c, e, f; Table 4). These results collectively demonstrated the added value of the constructed model to the existing clinical models.

The WSI-based score refines patient selection for adjuvant chemotherapy
We then investigated whether our model could help refine subgroups of patients who could mostly benefit from adjuvant chemotherapy. In validation cohort 1, neither stage IB ($p = 0.551$), stage II ($p = 0.116$), nor stage III patients ($p = 0.068$) significantly benefited from adjuvant chemotherapy (Supplementary Fig. 2). Further analysis with combination of the constructed model revealed no survival benefit for patients in the low-risk groups of patients with stage IB ($p = 0.974$, Fig. 4a), stage II ($p = 0.800$, Fig. 4b), or stage III ($p = 0.464$, Fig. 4c) disease. For patients in the high-risk groups, a survival advantage was acquired for patients in stage III ($p = 0.030$, Fig. 4f) and potentially acquired for patients in stage II ($p = 0.077$, Fig. 4e), but no survival benefit was observed for patients in stage IB ($p = 0.367$, Fig. 4d). Similar results were obtained for validation cohort 2 (Supplementary Fig. 2; Fig. 4g–l).

Interpretation of the deep learning model
To better understand the pathological mechanism underlying this prediction, we used heatmap visualization to explore the pathomorphological features of our model. As illustrated in Fig. 1d, micropapillary components were identified in the 'high-risk' region of the patient with stage IA tumor. Moreover, acinar patterns with tertiary lymphoid structures were characterized in the 'low-risk' region of the patient with stage III tumor. This reflects the substantial associations of the constructed model with the current pathological factors and its ability to serve as a complementary biomarker.

Patients in different risk groups present significant heterogeneity in gene expression patterns (Fig. 5a). In GO analyses (Fig. 5b), patients with the model defined high-risk group were associated with pathways representing tumor metabolism and proliferation such as cellular metabolic process, protein metabolic process, cellular component organization, and cellular component organization or biogenesis. Furthermore, as shown in Fig. 5c, tumors in two groups were characterized by diverse immune infiltration patterns. According to results of ssGSEA (Fig. 5d), patients with the model defined high-risk group yielded significantly less infiltrations of activated CD4 T cell, activated dendritic cell, central memory CD4 T cell, central memory CD8 T cell, effector memeory CD4 T cell, immature B cell, immature dendritic cell, macrophage, MDSC, natural killer T cell, and T follicular helper cell.

Discussion
Recently, the development of digital pathology has provided important information for precise risk stratification and treatment planning. However, predicting prognosis (time-to-event) is considered a more sophisticated problem than a conventional regression task, mainly due to the fact that some patients have not experienced the expected outcomes (death, recurrence, etc.). For this reason, the number of studies using WSIs for predicting prognosis is relatively small[9,10,14–17]. From the perspective of technology, these studies have mostly used convolution neural network[9,10,15–17] to automatically extract features. However, WSIs have many pixels containing lots of invalid information, which seriously affects the accuracy and efficiency of the convolutional neural network. Lee et al.[14] used the aggregation algorithm
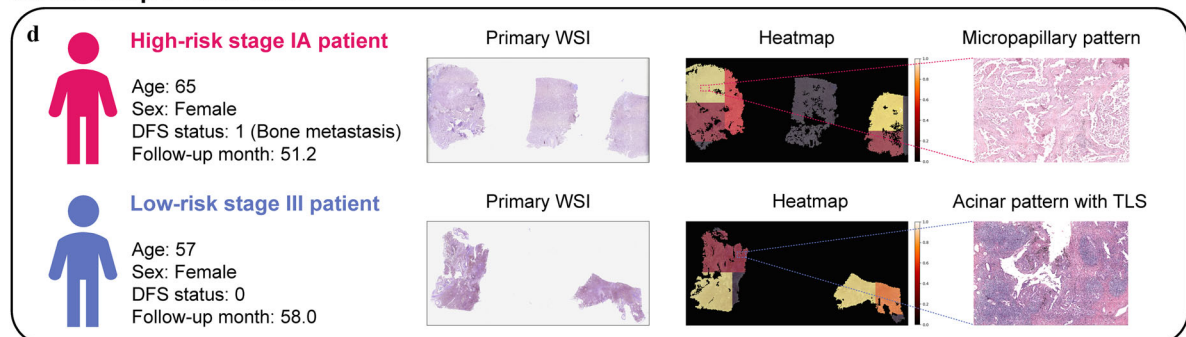
I. Model Construction
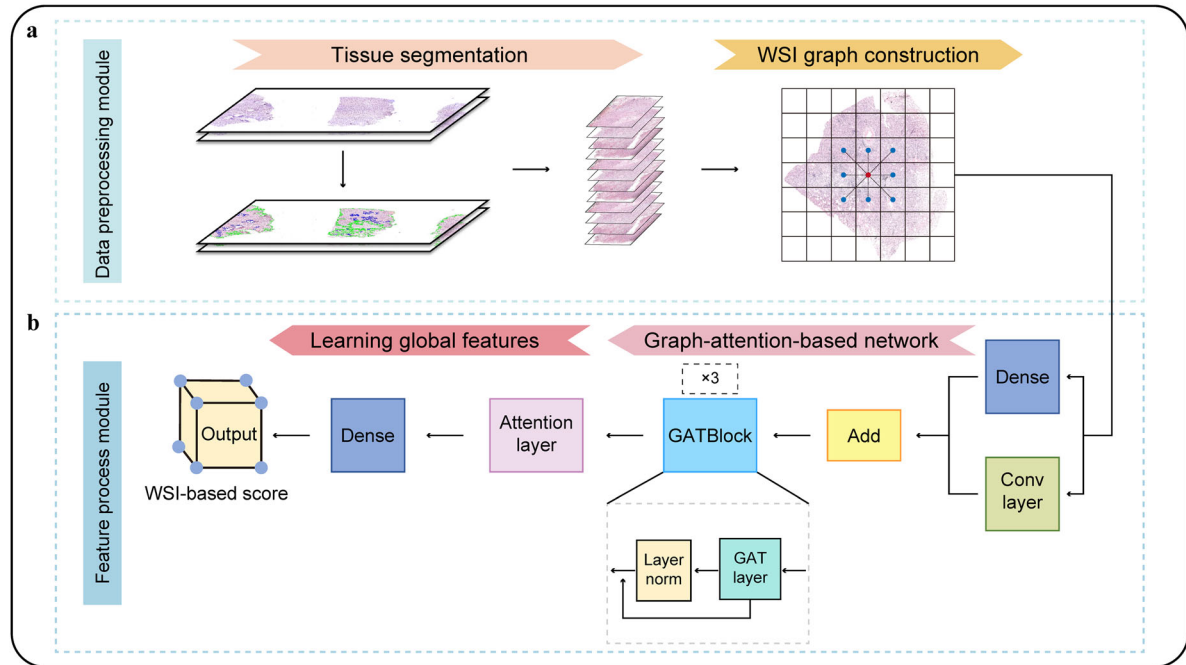
II. Efficiency Validation
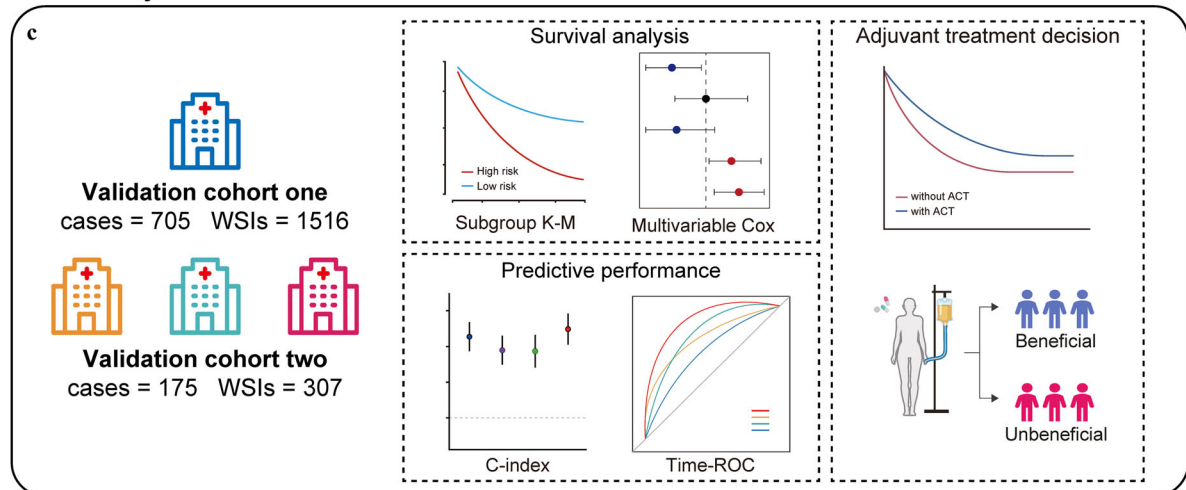
III. Heatmap Visualization

Fig. 1 | Schematic illustration of the overall study design. a, **b** Model construction; **c** efficiency validation; **d** heatmap visualization. WSI whole slide image, ROC receiver operating characteristic, ACT adjuvant chemotherapy, TLS tertiary lymphoid structure.

to obtain the superpatch graph before using GNN, which bring a great impact on the working efficiency of the prediction system. Further, Mobadersany et al.[16] used region of interest from WSI as input information, which increased the workload of clinicians. Compared with the methods mentioned above, the proposed model c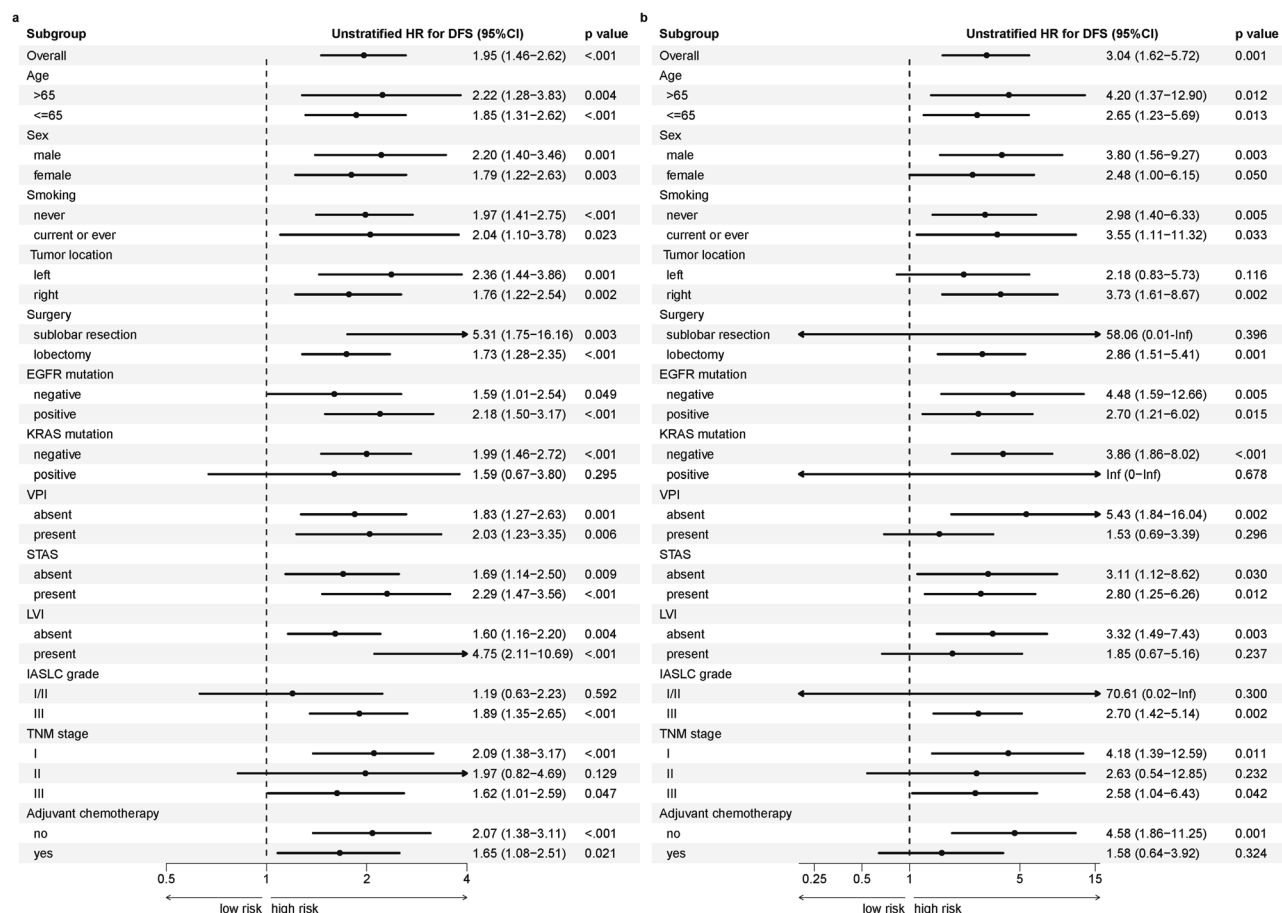onverts WSI into graph-based data and introduces the attention mechanism to assign different weights to different nodes, which effectively reduces the computational complexity and improves the prediction accuracy. Moreover, the proposed model was established without the facility of pathologists, overcoming the shortcoming that deep learning algorithms rely on manual annotations and the expertise

Table 1 | Baseline characteristics of patients in validation cohort 1 and validation cohort 2

Variables	Validation cohort 1 (*n* = 705)				Validation cohort 2 (*n* = 175)			
	Entire cohort	Low-risk group (*n* = 352)	High-risk group (*n* = 353)	*p*	Entire cohort	Low-risk group (*n* = 88)	High-risk group (*n* = 87)	*p*
Age, year, median (IQR)	61 (56–67)	62 (52–67)	61 (55–67)	0.122	61 (54–67)	61 (55–69)	59 (53–66)	0.547
>65	205 (29.1)	110 (31.3)	95 (26.9)	0.205	53 (30.3)	29 (33.0)	24 (27.6)	0.440
≤65	500 (70.9)	242 (68.8)	258 (73.1)		122 (69.7)	59 (67.0)	63 (72.4)	
Sex, *n* (%)				0.853				0.001
Male	298 (42.3)	150 (42.6)	148 (41.9)		84 (48.0)	53 (60.2)	31 (35.6)	
Female	407 (57.7)	202 (57.4)	205 (58.1)		91 (52.0)	35 (39.8)	56 (64.4)	
Smoking history, *n* (%)				0.083				0.154
Never	567 (80.4)	274 (77.8)	293 (83.0)		137 (78.3)	65 (73.7)	72 (82.8)	
Current or ever	138 (19.6)	78 (22.2)	60 (17.0)		38 (21.7)	23 (26.1)	15 (17.2)	
Tumor location, *n* (%)				0.329				0.954
Left lung	273 (38.7)	130 (36.9)	143 (40.5)		64 (36.6)	32 (36.4)	32 (36.8)	
Right lung	432 (61.3)	222 (63.1)	210 (59.5)		111 (63.4)	56 (63.6)	55 (63.2)	
Surgery type, *n* (%)				0.147				0.583
Sublobar resection	95 (13.5)	54 (15.3)	41 (11.6)		16 (9.1)	7 (8.0)	9 (10.3)	
Lobectomy or above	610 (86.5)	298 (84.7)	312 (88.4)		159 (90.9)	81 (92.0)	78 (89.7)	
EGFR mutation, *n* (%)				0.087				0.271
Positive	459 (65.1)	240 (68.2)	219 (62.0)		122 (69.7)	58 (65.9)	64 (73.6)	
Negative	246 (34.9)	112 (31.8)	134 (38.0)		53 (30.3)	30 (34.1)	23 (26.4)	
KRAS mutation, *n* (%)				0.591				0.107
Positive	58 (8.2)	27 (7.7)	31 (8.8)		10 (5.7)	8 (9.1)	2 (2.3)	
Negative	647 (91.8)	325 (92.3)	322 (91.2)		165 (94.3)	90 (90.9)	85 (97.7)	
VPI, *n* (%)				0.098				0.089
Present	169 (24.0)	75 (21.3)	94 (26.6)		52 (29.7)	21 (23.9)	31 (35.6)	
Absent	536 (76.0)	277 (78.7)	258 (73.4)		123 (70.3)	67 (76.1)	56 (64.4)	
STAS, *n* (%)				0.048				0.493
Present	231 (32.8)	103 (29.3)	128 (36.3)		64 (36.6)	30 (34.1)	34 (39.1)	
Absent	474 (67.2)	249 (70.7)	225 (63.7)		111 (63.4)	58 (65.9)	53 (60.9)	
LVI, *n* (%)				0.163				0.101
Present	99 (14.0)	43 (12.2)	56 (15.9)		30 (17.1)	11 (12.5)	19 (21.8)	
Absent	606 (86.0)	309 (87.8)	297 (84.1)		145 (82.9)	77 (87.5)	68 (78.2)	
IASLC grade, *n* (%)				<0.001				0.826
I	129 (18.3)	73 (20.7)	56 (15.9)		33 (18.9)	18 (20.5)	15 (17.2)	
II	199 (28.2)	119 (33.8)	80 (22.7)		45 (25.7)	23 (26.1)	22 (25.3)	
III	377 (53.5)	160 (45.5)	217 (61.5)		97 (55.4)	47 (53.4)	50 (57.5)	
TNM stage, *n* (%)				0.234				0.382
I	542 (76.9)	279 (79.3)	263 (74.5)		125 (71.4)	67 (76.1)	58 (66.7)	
II	57 (8.1)	28 (8.0)	29 (8.2)		19 (10.9)	8 (9.1)	11 (12.6)	
III	106 (15.0)	45 (12.8)	61 (17.3)		31 (17.7)	13 (14.8)	18 (20.7)	
Adjuvant chemotherapy, *n* (%)	247 (35.0)	107 (30.4)	140 (39.7)	0.010	56 (32.0)	26 (29.5)	30 (34.5)	0.484
Follow-up time, months, median (range)	68.0 (1.0–95.0)	65.0 (1.0–87.0)	69.0 (1.0–95.0)		68.0 (1.0–91.0)	61.0 (1.0–77.0)	72.0 (1.0–91.0)	
Recurrence, *n* (%)	200 (28.4)	68 (19.3)	132 (37.4)	<0.001	50 (28.6)	13 (14.8)	37 (42.5)	<0.001
WSI number per patient, median (range)	2 (1–4)	2 (1–4)	2 (1–4)	0.919	2 (1–4)	2 (1–4)	2 (1–4)	0.152
WSI number, total	1516	758	758		307	158	149	

EGFR epidermal growth factor receptor, *KRAS* Kirsten rat sarcoma viral oncogene, *VPI* visceral pleural invasion, *STAS* spread through air space, *LVI* lymph-vascular invasion, *IASLC* International Association for the Study of Lung Cancer, *WSI* whole slide image, *IQR* interquartile range, *HR* hazard ratio, *CI* confidence interval.

a — Validation cohort 1

Subgroup	Unstratified HR for DFS (95%CI)	p value
Overall	1.95 (1.46–2.62)	<.001
Age		
>65	2.22 (1.28–3.83)	0.004
<=65	1.85 (1.31–2.62)	<.001
Sex		
male	2.20 (1.40–3.46)	0.001
female	1.79 (1.22–2.63)	0.003
Smoking		
never	1.97 (1.41–2.75)	<.001
current or ever	2.04 (1.10–3.78)	0.023
Tumor location		
left	2.36 (1.44–3.86)	0.001
right	1.76 (1.22–2.54)	0.002
Surgery		
sublobar resection	5.31 (1.75–16.16)	0.003
lobectomy	1.73 (1.28–2.35)	<.001
EGFR mutation		
negative	1.59 (1.01–2.54)	0.049
positive	2.18 (1.50–3.17)	<.001
KRAS mutation		
negative	1.99 (1.46–2.72)	<.001
positive	1.59 (0.67–3.80)	0.295
VPI		
absent	1.83 (1.27–2.63)	0.001
present	2.03 (1.23–3.35)	0.006
STAS		
absent	1.69 (1.14–2.50)	0.009
present	2.29 (1.47–3.56)	<.001
LVI		
absent	1.60 (1.16–2.20)	0.004
present	4.75 (2.11–10.69)	<.001
IASLC grade		
I/II	1.19 (0.63–2.23)	0.592
III	1.89 (1.35–2.65)	<.001
TNM stage		
I	2.09 (1.38–3.17)	<.001
II	1.97 (0.82–4.69)	0.129
III	1.62 (1.01–2.59)	0.047
Adjuvant chemotherapy		
no	2.07 (1.38–3.11)	<.001
yes	1.65 (1.08–2.51)	0.021

Axis: 0.5 1 2 4 — low risk ← → high risk

b — Validation cohort 2

Subgroup	Unstratified HR for DFS (95%CI)	p value
Overall	3.04 (1.62–5.72)	0.001
Age		
>65	4.20 (1.37–12.90)	0.012
<=65	2.65 (1.23–5.69)	0.013
Sex		
male	3.80 (1.56–9.27)	0.003
female	2.48 (1.00–6.15)	0.050
Smoking		
never	2.98 (1.40–6.33)	0.005
current or ever	3.55 (1.11–11.32)	0.033
Tumor location		
left	2.18 (0.83–5.73)	0.116
right	3.73 (1.61–8.67)	0.002
Surgery		
sublobar resection	58.06 (0.01-Inf)	0.396
lobectomy	2.86 (1.51–5.41)	0.001
EGFR mutation		
negative	4.48 (1.59–12.66)	0.005
positive	2.70 (1.21–6.02)	0.015
KRAS mutation		
negative	3.86 (1.86–8.02)	<.001
positive	Inf (0-Inf)	0.678
VPI		
absent	5.43 (1.84–16.04)	0.002
present	1.53 (0.69–3.39)	0.296
STAS		
absent	3.11 (1.12–8.62)	0.030
present	2.80 (1.25–6.26)	0.012
LVI		
absent	3.32 (1.49–7.43)	0.003
present	1.85 (0.67–5.16)	0.237
IASLC grade		
I/II	70.61 (0.02-Inf)	0.300
III	2.70 (1.42–5.14)	0.002
TNM stage		
I	4.18 (1.39–12.59)	0.011
II	2.63 (0.54–12.85)	0.232
III	2.58 (1.04–6.43)	0.042
Adjuvant chemotherapy		
no	4.58 (1.86–11.25)	0.001
yes	1.58 (0.64–3.92)	0.324

Axis: 0.25 0.5 1 5 15 — low risk ← → high risk

Fig. 2 | Analysis of DFS in specified subgroups. a Validation cohort 1; **b** validation cohort 2. EGFR epidermal growth factor receptor, KRAS Kirsten rat sarcoma viral oncogene, VPI visceral pleural invasion, STAS spread through air space, LVI lymph- vascular invasion, IASLC International Association for the Study of Lung Cancer, HR hazard ratio, CI confidence interval, DFS disease-free survival.

of pathologists to a certain extent[18,19], which may help to improve the generalization of the model.

On the one hand, although the predictive performance for prognosis of our model did not outperform some of the pathological factors, it remained statistically significant in the multivariable analysis when it was combined with TNM stage and IASLC grade, indicating the added value of the model and its ability to serve as a complementary biomarker for survival stratification. On the other hand, the presence of high-risk pathological factors, including VPI, STAS, LVI, and high-grade tumors, was more common in the model-defined high-risk groups, suggesting substantial associations between the constructed model and the current well-defined pathological factors.

A large meta-analysis revealed that adjuvant chemotherapy could yield an overall survival (OS) benefit of 5% at 5 year, however, the statistically significance was not reached (HR 0.87, $p = 0.08$)[20]. Following this study, randomized trials evaluating the efficiency of adjuvant chemotherapy were conducted for a decade. Some of the large trials successfully demonstrated the OS benefit[21,22], while others failed[23]. Afterwards, the Lung Adjuvant Cisplatin Evaluation (LACE) study further confirmed the effect of adjuvant chemotherapy on both OS and DFS[24]. Nevertheless, we must be clear that the overall benefit from adjuvant chemotherapy is limited: stage II-III patients may mostly benefit, stage IB patients may only have trend toward benefit, while stage IA patients may experience deleterious effect. We need to identify subgroups of patients who may particularly benefit from adjuvant chemotherapy. In the current study, no significant survival benefits from adjuvant chemotherapy were acquired across the overall population of patients with stage IB, stage II, or stage III disease. However, combining TNM stage with our constructed model helps to distinguish a survival advantage for high-risk stage III patients, and a potential survival advantage for high-risk stage II patients (statistical significance was not reached for this group perhaps for the limitation of the relatively small sample size). According to our results, we advocate adjuvant chemotherapy for high-risk stage II-III patients and to avoid unnecessary chemotherapy for other patients.

The results demonstrated that our established model exhibits significant biological relevance. The model outputs are likely associated with genes and molecular pathways that promote tumor proliferation, and high-risk patients show significantly lower levels of immune cell infiltration. This partly explains the model's predictive capability for prognosis and adjuvant chemotherapy decision-making in lung cancer patients.

Despite the promising results obtained in the present study, several limitations should be declared. Firstly, the retrospectively nature of the study may limit the statistical power and hinder the generalization of the results to other centers and regions, especially the results regarding adjuvant therapy, prospective validation with larger sample size is warranted. Second, although our model could be used as a complement to the existing prognostic models of lung adenocarcinoma, there remains much room for improvement in its ability to predict prognosis. Multiomics data integrating radiology, pathology, molecular, and other modalities are needed to establish more efficient and robust models in the future.

In summary, our constructed model can predict the recurrence risk of resected lung adenocarcinoma without the need for annotations from pathologists, which can complement the current prognostic models.

Table 2 | Multivariable Cox analysis of disease-free survival in validation cohort 1

| Variables | Validation cohort 1 (n = 705) | | | | | |
| | Univariable | | Multivariable[a] | | Multivariable[b] | |
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Age (continuous)	1.00 (0.98–1.01)	0.417				
Sex (male vs. female)	1.10 (0.82–1.44)	0.551				
Smoke (yes vs. no)	1.21 (0.86–1.70)	0.275				
Location (left vs. right)	0.92 (0.70–1.23)	0.586				
Surgery (lobectomy vs. sublobar resection)	1.61 (0.99–2.62)	0.054	0.76 (0.46–1.25)	0.279	0.75 (0.46–1.23)	0.258
EGFR mutation (positive vs. negative)	0.81 (0.61–1.08)	0.151				
KRAS mutation (positive vs. negative)	1.31 (0.81–2.13)	0.277				
VPI (present vs. absent)	2.17 (1.63–2.90)	<0.001	1.12 (0.82–1.53)	0.466	1.09 (0.80–1.49)	0.575
STAS (present vs. absent)	2.09 (1.58–2.76)	<0.001	0.99 (0.72–1.34)	0.922	1.01 (0.74–1.37)	0.965
LVI (present vs. absent)	1.96 (1.40–2.75)	<0.001	0.96 (0.67–1.37)	0.825	0.98 (0.69–1.40)	0.918
IASLC grade						
II vs. I	14.13 (3.41–58.65)	<0.001	13.17 (3.16–54.96)	<0.001	13.54 (3.24–56.52)	<0.001
III vs. I	37.77 (9.36–152.50)	<0.001	26.99 (6.50–112.13)	<0.001	25.69 (6.19–106.63)	<0.001
TNM stage						
II vs. I	2.25 (1.42–3.57)	0.001	1.59 (0.98–2.57)	0.061	1.67 (1.03–2.71)	0.037
III vs. I	5.83 (4.31–7.90)	<0.001	4.04 (2.83–5.76)	<0.001	4.22 (2.95–6.05)	<0.001
Adjuvant chemotherapy (yes vs. no)	1.91 (1.45–2.52)	<0.001	0.82 (0.60–1.13)	0.219	0.77 (0.56–1.07)	0.117
WSI-based score (high risk vs. low risk)	1.95 (1.46–2.62)	<0.001			1.82 (1.35–2.44)	<0.001

EGFR epidermal growth factor receptor, *KRAS* Kirsten rat sarcoma viral oncogene, *VPI* visceral pleural invasion, *STAS* spread through air space, *LVI* lymph-vascular invasion, *IASLC* International Association for the Study of Lung Cancer, *WSI* whole slide image, *HR* hazard ratio, *CI* confidence interval, *DFS* disease-free survival.
[a]Multivariable model without incorporating the model defined low- and high-risk groups.
[b]Multivariable model with incorporating the model defined low- and high-risk groups.

Moreover, the model defined high- and low-risk groups may help to guide adjuvant therapy strategies in clinical practice.

Methods

Participants and study design

This multicenter study was approved by the Ethics Committee and Institutional Review Board of Shanghai Pulmonary Hospital (No. K23-292), the First Affiliated Hospital of Nanchang University, Ningbo Hwamei Hospital, the First Affiliated Hospital of Lanzhou University, and followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement[25] (Supplementary Note 1). The informed consent was waived as this was a retrospective study.

We used 3712 H&E-stained, formalin-fixed and paraffin-embedded (FFPE) tumor tissue sections from 1705 patients with surgically resected lung adenocarcinoma. Patients with stage I-III disease and available clinicopathological data and follow-up information were included. Patients with stage IV disease, a history of neoadjuvant therapy, and no available follow-up information or tumor tissue sections were excluded. For each patient in the training and validation sets, digital WSIs were scanned from the corresponding H&E-stained tumor tissue sections.

To train the model, we used a dataset of 1889 sections from 825 patients who underwent surgery at Shanghai Pulmonary Hospital between January 2012 and December 2012 (Supplementary Table 2). We applied deep learning techniques to develop a histopathological model, the patient-level WSI-based score, to predict the disease-free survival (DFS) of patients with resected lung adenocarcinoma. We then evaluated the capability of our model for survival stratification and investigated whether it could help refine the populations of patients likely to benefit from adjuvant chemotherapy in two separate validation sets. Finally, WSI heatmaps were obtained to explore the pathological features underlying the predictions (see Fig. 1 for the study design). In addition, the underlying biological basis of the model was also explored to enhance the interpretability of the model. The validation cohort 1 included 1516 sections from 705 patients who underwent surgery at Shanghai Pulmonary Hospital between January 2015 and June 2015. The validation cohort 2 included 307 sections from 175 patients between January 2015 and December 2015 from three departments of thoracic surgery: the First Affiliated Hospital of Nanchang University, Ningbo Hwamei Hospital, and the First Affiliated Hospital of Lanzhou University.

Clinical data, including age, sex, smoking history, tumor location, surgery type, and TNM stage, were available for both the training and validation sets. To compare the performance of the constructed model in prognosis prediction with that of the current clinical models, the VPI, STAS, and LVI status and IASLC tumor grade[26] were re-evaluated by two of our experienced pathologists (C.W., L.H.) for patients in the validation sets.

Four 21-day cycles of intravenous chemotherapy of cisplatin 75 mg/m^2 or carboplatin AUC 5 on day 1 plus pemetrexed 500 mg/m^2 on day 1 were administrated after thorough evaluation of the patients' conditions and discussion among a group of surgeons and oncologists at our centers.

WSI-based score for recurrence risk prediction

The patient-level survival prediction model in this paper is a multiple-classification model based on variable length input. Since the number of WSIs obtained for each patient varied, and the effective area of different WSIs also varied considerably, we need to utilize a model that can handle inputs of variable length. In addition, since hundreds of millions of pixels are contained in WSIs, efficient compression of the input data was also considered to be crucial. Thus, this paper presents a graph-attention-based multiple-instance neural network (GAMINN) for processing variable WSIs for survival prediction. The total analysis system contains the following modules, whose details are shown below.

Table 3 | Multivariable Cox analysis of disease-free survival in validation cohort 2

Variables	Validation cohort 2 (n = 175)					
	Univariable		Multivariable[a]		Multivariable[b]	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Age (continuous)	1.01 (0.99–1.05)	0.347				
Sex (male vs. female)	0.92 (0.53–1.61)	0.777				
Smoke (yes vs. no)	1.55 (0.84–2.88)	0.164				
Location (left vs. right)	1.06 (0.60–1.88)	0.833				
Surgery (lobectomy vs. sublobar resection)	1.40 (0.44–4.51)	0.571				
EGFR mutation (positive vs. negative)	0.70 (0.39–1.24)	0.218				
KRAS mutation (positive vs. negative)	1.75 (0.69–4.41)	0.237				
VPI (present vs. absent)	3.91 (2.23–6.84)	<0.001	1.59 (0.86–2.95)	0.138	1.44 (0.79–2.62)	0.238
STAS (present vs. absent)	3.79 (2.13–6.75)	<0.001	1.30 (0.69–2.45)	0.421	1.15 (0.60–2.21)	0.666
LVI (present vs. absent)	3.90 (2.20–6.93)	<0.001	1.66 (0.90–3.05)	0.106	1.48 (0.80–2.74)	0.214
IASLC grade						
II vs. I	2.21 (0.23–21.23)	0.493	1.94 (0.20–18.67)	0.568	1.88 (0.19–18.13)	0.587
III vs. I	22.22 (3.06–161.32)	0.002	9.02 (1.13–72.27)	0.038	10.29 (1.28–82.93)	0.028
TNM stage						
II vs. I	3.63 (1.64–8.04)	0.002	1.66 (0.71–3.87)	0.244	1.54 (0.67–3.56)	0.314
III vs. I	7.14 (3.80–13.40)	<0.001	2.73 (1.36–5.51)	0.005	3.08 (1.53–6.21)	0.002
Adjuvant chemotherapy (yes vs. no)	1.78 (1.01–3.12)	0.046	0.76 (0.42–1.40)	0.385	0.69 (0.38–1.25)	0.217
WSI-based score (high risk vs. low risk)	3.04 (1.62–5.72)	0.001			2.96 (1.55–5.66)	0.001

EGFR epidermal growth factor receptor, *KRAS* Kirsten rat sarcoma viral oncogene, *VPI* visceral pleural invasion, *STAS* spread through air space, *LVI* lymph-vascular invasion, *IASLC* International Association for the Study of Lung Cancer, *WSI* whole slide image, *HR* hazard ratio, *CI* confidence interval, *DFS* disease-free survival.
[a]Multivariable model without incorporating the model defined low- and high-risk groups.
[b]Multivariable model with incorporating the model defined low- and high-risk groups.

Data preprocessing, tissue segmentation and feature compression

Because each WSI contains a large number of pixels, it is difficult for a deep learning model to directly process the WSI and obtain good prediction results. Moreover, each WSI contains a large amount of invalid interference information, which not only affects the subsequent analysis performance but also consumes computing resources. Therefore, in this paper, the CLAM model[27] was used to classify the tissue regions of pathological images, which can efficiently and accurately classify the regions with high diagnostic value based on low computational burden. Figure 1a shows the tissue profile extracted by the CLAM model, which was used to segment the tissue region effectively and eliminate interference from the invalid region. Thereafter, we partitioned the extracted regions into patch slices, each with a size of 256×256 pixels. Meanwhile, we used the ResNet50 model pretrained on the ImageNet2012 dataset to process the extracted patch slices and extract the morphological features of each slice, where the morphological feature dimension of each slice was 1024.

WSI graph construction

For each patch, we saved the position coordinates of each patch in the pathological image from the tissue segmentation and built an adjacency matrix A_j using the fast approximation KNN ($k = 8$). The adjacency matrix A_j modeled a 3×3 image receptive field in the pathological image. Finally, we combined the feature matrix X_j and adjacency matrix A_j to construct the subgraph $G_j = (X_j, A_j)$, and all WSIs for the same patient were constructed as $G = \{G_j\}_{j=1}$. Figure 1a shows the process of WSI graph construction.

Feature process module

We combined a graph convolutional neural network and a self-attention mechanism to process the feature input of the graph structure, effectively grasp the implicit relationships between slices, and obtain an effective representation for prognostic risk assessment. Moreover, according to the different numbers of patient-level input WSIs, a multi-instance pooling method was designed to effectively obtain the features of different WSIs of the same patient, thereby improving the final prognosis prediction performance.

Graph-attention-based network

To better handle the input data of the graph structure, we use graph neural network to automatically extract features. Similar to convolutional neural networks, graph convolutional neural networks (GCNs) have powerful feature learning capabilities, in which the convolution of a certain point can be viewed as a weighted sum of the neighbors of the point. However, the GCN treats all neighboring nodes equally during convolution and cannot assign different weights according to the importance of the nodes. Assuming that there are N nodes in a graph, in practical analysis, the contributions of different adjacent nodes to the target node should also be different. To better distribute weights among different nodes, we used an attention mechanism to uniformly normalize the correlation calculated between the target node and all its neighbors.

$$a_{i,j} = \frac{\exp(LeakyReLU(a^T[WX_i \parallel WX_j]))}{\sum_{k \in N_i} \exp(LeakyReLU(a^T[WX_i \parallel WX_k]))} \tag{1}$$

$\parallel$ is the concatenation operation, W is the linear transformation matrix, a is a renewable matrix, and $a_{i,j}$ are the connection degrees of node j to node i. By combining the GCN with then attention mechanism, we constructed a graph-attention-based (GAT) layer for subsequent analysis.

Learning global features

We build an end-to-end differentiable function F_{GAT}, using a GAT layer to mine the node features of each neighbor in the space. To further learn
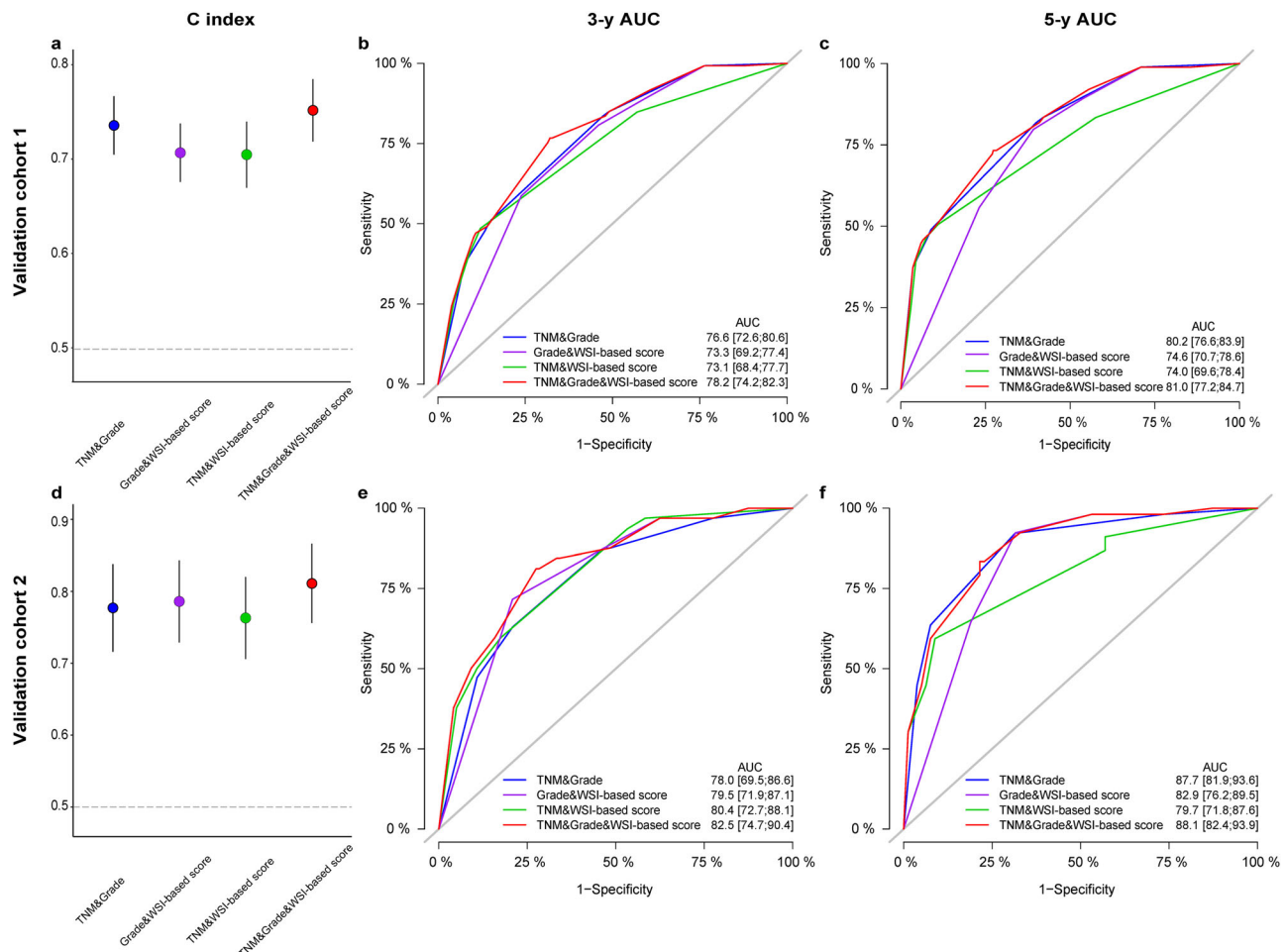
Fig. 3 | C-index and time-dependent ROC curves at 3 and 5 years for the combination of variables independently predicting DFS according to multivariable Cox regression analysis. a–c Validation cohort 1; **d–f** validation cohort 2. ROC receiver operating characteristic, AUC area under the curve, DFS disease-free survival.

Table 4 | Performance of the constructed model combined with existing pathological variables according to multivariable analysis

	C index	p	3-y AUC	p	5-y AUC	p
Validation cohort 1						
IASLC grade & TNM stage	0.737 (0.706–0.768)	<0.001	0.766 (0.726–0.806)	<0.001	0.802 (0.766–0.839)	<0.001
WSI-based score & IASLC grade	0.708 (0.677–0.739)	<0.001	0.733 (0.692–0.774)	<0.001	0.746 (0.707–0.786)	<.001
WSI-based score & TNM stage	0.706 (0.671–0.741)	<0.001	0.731 (0.684–0.777)	<0.001	0.740 (0.696–0.784)	<0.001
WSI-based score & IASLC grade & TNM stage	0.753 (0.720–0.786)	–	0.782 (0.742–0.823)	–	0.810 (0.772–0.847)	–
Validation cohort 2						
IASLC grade & TNM stage	0.777 (0.716–0.838)	<0.001	0.780 (0.695–0.866)	<0.001	0.877 (0.819–0.936)	<0.001
WSI-based score & IASLC grade	0.786 (0.729–0.843)	<0.001	0.795 (0.719–0.871)	<0.001	0.829 (0.762–0.895)	<0.001
WSI-based score & TNM stage	0.763 (0.706–0.820)	<0.001	0.804 (0.727–0.881)	<0.001	0.797 (0.718–0.876)	<0.001
WSI-based score & IASLC grade & TNM stage	0.811 (0.756–0.866)	–	0.825 (0.747–0.904)	–	0.881 (0.824–0.939)	–

The p values refer to the comparison between the combination of two independent predictors and the combination of three independent predictors in the multivariable Cox regression model.
AUC area under the curve, *IASLC* International Association for the Study of Lung Cancer.

the global morphological features of pathological images and avoid gradient vanishing in the network, inspired by the idea of residual learning, we used $F_{GAT}(l)$ as a residual map, which allows the superposition of multiple layers of $F_{GAT}(l)$ together, where the output of F_{GAT} (l) is added to the input.

$$G^{l+1} = F^l_{GAT}(G^l; \Phi^l, \rho^l, \zeta^l) + G^l \qquad (2)$$
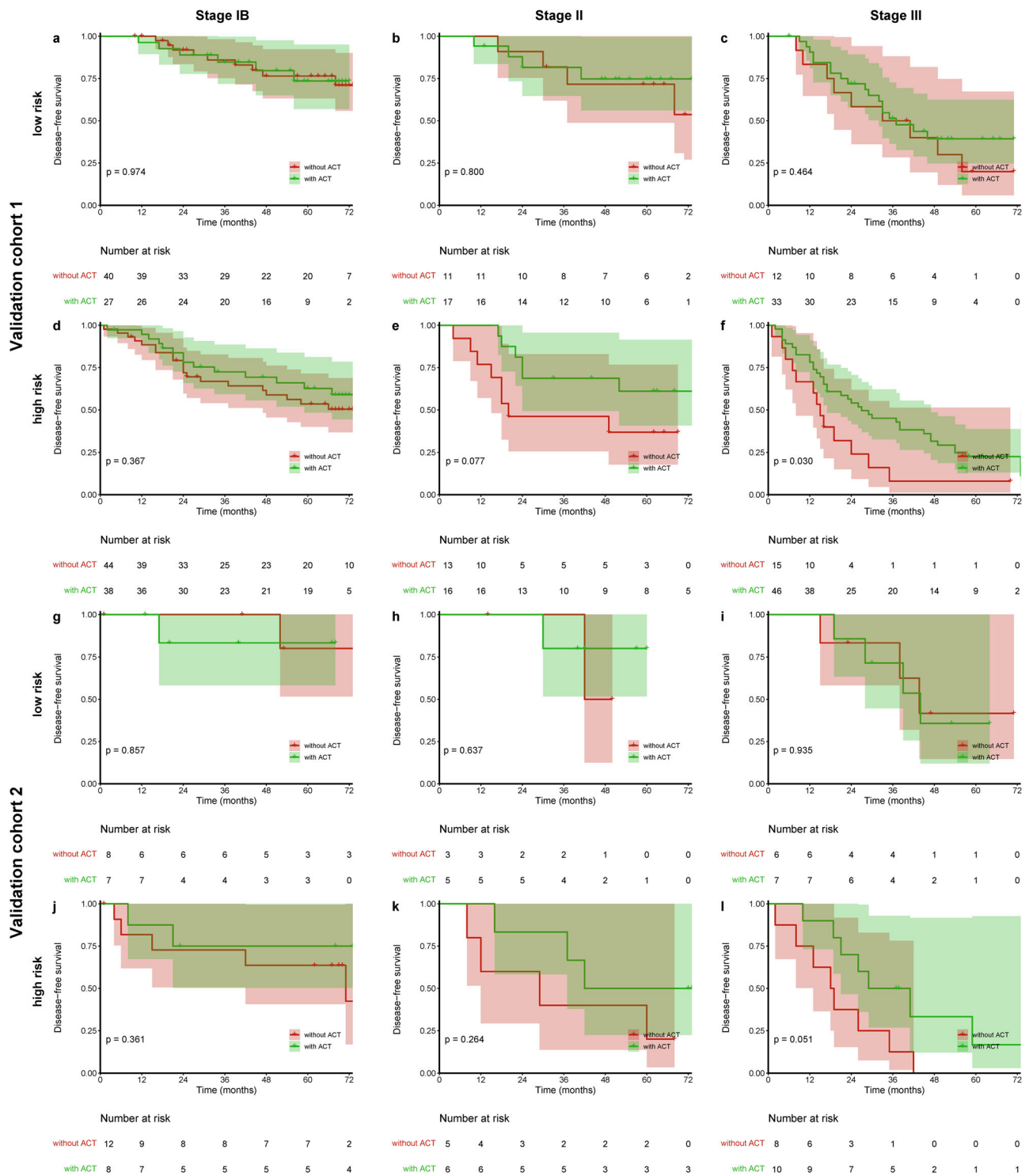
Fig. 4 | The combination of TNM stage and the constructed model helps to identify patient subgroups who may particularly benefit from postoperative chemotherapy.
a–f Validation cohort 1; **g–l** validation cohort 2. ACT adjuvant chemotherapy.

where φ^l is a message construction function that calculates the association characteristics between node u and its neighbor node v, ρ^l is an aggregation function that aggregates all the features passed to v, and ζ^l is an update function that updates the existing node features at node v with the aggregated features X^{l+1}_v.

$$m^l_v = \rho^l(\Phi^l(X^l_v, X^l_u)) \tag{3}$$

$$X^{l+1}_v = \zeta^l(X^l_v, m^l_v) \tag{4}$$

We implemented the main model structure of GAMINN using a 3-layer residual GAT model. In addition, we output the last GAT layer to the fully connected layer and aggregated the different WSI features in the same patient to achieve better patient-level feature expression (Fig. 1b).
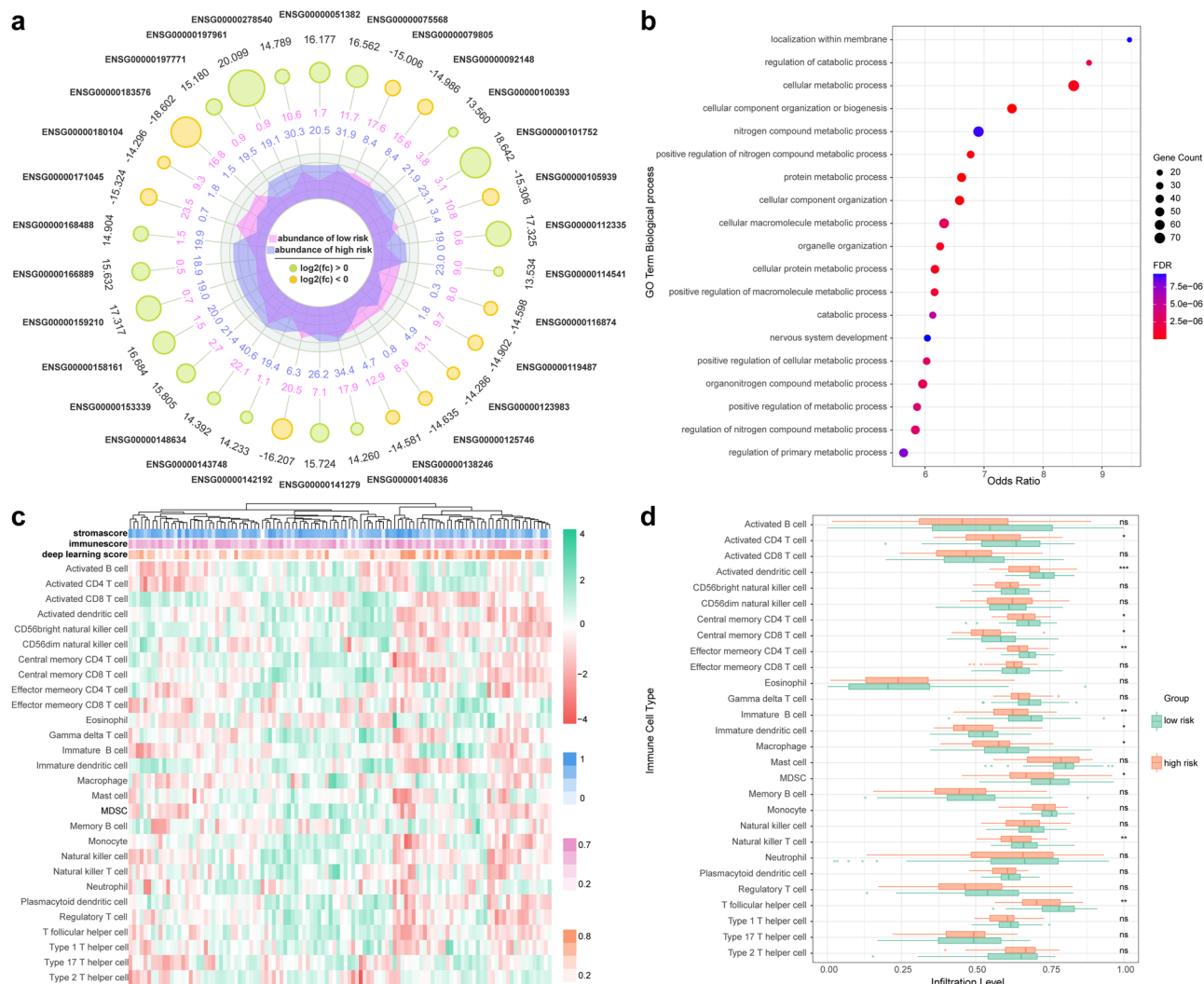
Fig. 5 | Biological basis of the deep learning model. a Radar charts illustrating top 30 differential genes between low-risk and high-risk patients. **b** Dot plots showing the top 20 upregulated molecular pathways in high risk patients; **c** Heat map illustrating immune infiltration patterns between low-risk and high-risk patients; **d** Boxplots comparing proportions of infiltrated immune cells between low-risk and high-risk patients. GO gene ontology, FDR false discovery rate, MDSC myeloid-derived suppressor cells.

Details on network training

We use NVIDIA GeForce 3070 GPU RTX for training the model, which has 16 GB of memory. Additionally, we use the PyTorch library version 1.12.1 for training and evaluation. Adam optimizer is selected as the model optimizer, whose initial weight is 0.0002, and each batch contains a multiple pathological image data of patients. During the training process, the model is trained through 100 epochs and utilize Cox likelihood function as loss function, which is listed as follows:

$$Loss = -\sum_i \left(h_\theta(x_i) - \log \sum_{j \in R(T_i)} e^{h_\theta(x_j)} \right) \quad (5)$$

x_i represents ith cases, $h_\theta(.)$ means the risk score from the proposed model, and $R(T_i)$ is the list of patients with shorter survival time than the ith patient.

Biological basis of deep learning model

RNA-sequencing was performed in 112 patients in validation cohort 1, the TruSeq RNA Access Library Prep Kit (Illumina) was utilized to generate library and the paired-end sequencing based on an Illumina Novaseq™ 6000 was subsequently conducted. Among them, 63 patients were classified as low-risk and 49 as high-risk. We used the edgeR package to determine differentially expressed genes between two groups with standard of log fold changes more

than 1 and adjusted p values less than 0.05. Subsequently, Gene Ontology (GO) pathway analyses was performed to determine pathways related to the model defined risk groups. Additionally, the single sample gene set enrichment analysis (ssGSEA) was conducted with the GSVA package to quantify the relative infiltration of immune cell types in the tumor microenvironment.

Statistical analysis

DFS was defined as the time from surgery to the first-confirmed event of lung cancer recurrence. The Kaplan–Meier method and log-rank test were used to compare survival outcomes between groups. Cox regression analysis was performed to identify independent predictors of survival. The predictive performance of each model was assessed via the Harrell concordance index (C-index), time-dependent receiver operating characteristic (ROC) curves, and area under the curve (AUC) values at 3 and 5 years. The missing information was dealt with using the single imputation method. Statistical analysis was performed with R software (version 4.3.1). A two-sided p value less than 0.05 was considered to indicate statistical significance.

Data availability

The datasets analyzed in the current study are not publicly available due to patient privacy purposes, but are available upon reasonable request to the corresponding author. Access to the data will be restricted to non-commercial research.

Code availability

References

1. Chaft, J. E., Shyr, Y., Sepesi, B. & Forde, P. M. Preoperative and postoperative systemic therapy for operable non-small-cell lung cancer. *J. Clin. Oncol.* **40**, 546–555 (2022).
2. 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).
3. Zhong, Y. et al. Prognostic impact of tumour spread through air space in radiological subsolid and pure solid lung adenocarcinoma. *Eur. J. Cardiothorac. Surg.* **59**, 624–632 (2021).
4. Okiror, L. et al. Prognostic factors including lymphovascular invasion on survival for resected non-small cell lung cancer. *J. Thorac. Cardiovasc. Surg.* **156**, 785–793 (2018).
5. Fujikawa, R. et al. Clinicopathologic and genotypic features of lung adenocarcinoma characterized by the International Association for the Study of Lung Cancer Grading System. *J. Thorac. Oncol.* **17**, 700–707 (2022).
6. Hou, L. et al. Prognostic and predictive value of the newly proposed grading system of invasive pulmonary adenocarcinoma in Chinese patients: a retrospective multicohort study. *Mod. Pathol.* **35**, 749–756 (2022).
7. Detterbeck, F. C., Boffa, D. J., Kim, A. W. & Tanoue, L. T. The eighth edition lung cancer stage classification. *Chest* **151**, 193–203 (2017).
8. Wulczyn, E. et al. Interpretable survival prediction for colorectal cancer using deep learning. *NPJ Digit Med.* **4**. https://doi.org/10.1038/s41746-021-00427-2 (2021)
9. Qaiser, T. et al. Usability of deep learning and H&E images predict disease outcome-emerging tool to optimize clinical trials. *NPJ Precis. Oncol.* **6**. https://doi.org/10.1038/s41698-022-00275-7 (2022)
10. Courtiol, P. et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat. Med.* **25**, 1519–1525 (2019).
11. Ding, R. et al. Image analysis reveals molecularly distinct patterns of TILs in NSCLC associated with treatment outcome. *NPJ Precis. Oncol.* **6**. https://doi.org/10.1038/s41698-022-00277-5 (2022)
12. Rakaee, M. et al. Machine learning-based immune phenotypes correlate with STK11/KEAP1 co-mutations and prognosis in resectable NSCLC: a sub-study of the TNM-I trial. *Ann. Oncol.* **34**, 578–588 (2023).
13. Shi, J.-Y. et al. Exploring prognostic indicators in the pathological images of hepatocellular carcinoma based on deep learning. *Gut* **70**, 951–961 (2021).
14. Lee, Y. et al. Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning. *Nat. Biomed. Eng.* https://doi.org/10.1038/s41551-022-00923-0 (2022)
15. Jiang, X. et al. End-to-end prognostication in colorectal cancer by deep learning: a retrospective, multicentre study. *Lancet Digit Health* **6**, e33–e43 (2024).
16. Mobadersany, P. et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl Acad. Sci. USA* **115**, E2970–e2979 (2018).
17. Saillard, C. et al. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. *Hepatology* **72**, 2000–2013 (2020).
18. Zhang, Y. et al. Histopathology images-based deep learning prediction of prognosis and therapeutic response in small cell lung cancer. *NPJ Digit Med.* **7**, 15 (2024).
19. Nagpal, K. et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digit Med.* **2**, 48 (2019).
20. Alberti, W. et al. Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomised clinical trials. Non-small Cell Lung Cancer Collaborative Group. *BMJ* **311**, 899–909 (1995).
21. Arriagada, R. et al. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. *N. Engl. J. Med.* **350**, 351–360 (2004).
22. Douillard, J. Y. et al. Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIA non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. *Lancet Oncol.* **7**, 719–727 (2006).
23. Waller, D. et al. Chemotherapy for patients with non-small cell lung cancer: the surgical setting of the Big Lung Trial. *Eur. J. Cardiothorac. Surg.* **26**, 173–182 (2004).
24. Pignon, J. P. et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. *J. Clin. Oncol.* **26**, 3552–3559 (2008).
25. Collins, G. S. et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **385**, e078378 (2024).
26. Moreira, A. L. et al. A grading system for invasive pulmonary adenocarcinoma: a proposal from the International Association for the Study of Lung Cancer Pathology Committee. *J. Thorac. Oncol.* **15**, 1599–1610 (2020).
27. Lu, M. Y. et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**, 555–570 (2021).

Author contributions

T.C., C.C., L.H. designed this study. T.C., Y.J., J.W., X.S. analyzed the data and wrote the first version of the manuscript. Y.J., J.S. built the deep learning model. J.W., J.D., Y.Z., M.Z., L.X., Y.S. collected the clinicopathological data. L.H., B.Y., M.Y., M.M. provided the H&E slides. L.H., C.W. reviewed the H&E slides in two validation sets. C.C., Y.J., L.H., Y.Z. conceived the project and edited the paper. T.C., Y.J., Y.Z. mainly revised the manuscript. All authors read and approved the final manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

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