Research article

Spatial distance between tumor and lymphocyte can predict the survival of patients with resectable lung adenocarcinoma

Xipeng Pan [a,1], Siyang Feng [a,1], Yumeng Wang [a,c,1], Jiale Chen [a], Huan Lin [b,c,d], Zimin Wang [a], Feihu Hou [a], Cheng Lu [a,b,c], Xin Chen [a,e], Zhenbing Liu [a], Zhenhui Li [f,**], Yanfen Cui [b,c,g,h,***], Zaiyi Liu [a,b,c,*]

[a] School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, 541004, China
[b] Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China
[c] Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangzhou, 510080, China
[d] School of Medicine, South China University of Technology, Guangzhou, 510006, China
[e] Department of Radiology, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, 510180, China
[f] Department of Radiology, The Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Yunnan Cancer Centre, Kunming, 650118, China
[g] Guangdong Cardiovascular Institute, Guangzhou, 510080, China
[h] Department of Radiology, Shanxi Province Cancer Hospital, Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences/ Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, 030013, China

ARTICLE INFO

ABSTRACT

Background and objective: Spatial interaction between tumor-infiltrating lymphocytes (TILs) and tumor cells is valuable in predicting the effectiveness of immune response and prognosis amongst patients with lung adenocarcinoma (LUAD). Recent evidence suggests that the spatial distance between tumor cells and lymphocytes also influences the immune responses, but the distance analysis based on Hematoxylin and Eosin (H&E) -stained whole-slide images (WSIs) remains insufficient. To address this issue, we aim to explore the relationship between distance and prognosis prediction of patients with LUAD in this study.

Methods: We recruited patients with resectable LUAD from three independent cohorts in this multi-center study. We proposed a simple but effective deep learning-driven workflow to automatically segment different cell types in the tumor region using the HoVer-Net model, and quantified the spatial distance (DIST) between tumor cells and lymphocytes based on H&E-stained WSIs. The association of DIST with disease-free survival (DFS) was explored in the discovery set (D1, n = 276) and the two validation sets (V1, n = 139; V2, n = 115).

Results: In multivariable analysis, the low DIST group was associated with significantly better DFS in the discovery set (D1, HR, 0.61; 95 % CI, 0.40–0.94; p = 0.027) and the two validation sets (V1,

---

* Corresponding author Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China.
** Corresponding author.
*** Corresponding author. Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China.
E-mail addresses: lizhenhui621@qq.com (Z. Li), yanfen210@126.com (Y. Cui), liuzaiyi@gdph.org.cn (Z. Liu).
[1] Contributed equally.

HR, 0.54; 95 % CI, 0.32–0.91; p = 0.022; V2, HR, 0.44; 95 % CI, 0.24–0.81; p = 0.009). By integrating the DIST with clinicopathological factors, the integrated model (full model) had better discrimination for DFS in the discovery set (C-index, D1, 0.745 vs. 0.723) and the two validation sets (V1, 0.621 vs. 0.596; V2, 0.671 vs. 0.650). Furthermore, the computerized DIST was associated with immune phenotypes such as immune-desert and inflamed phenotypes.

Conclusions: The integration of DIST with clinicopathological factors could improve the stratification performance of patients with resectable LUAD, was beneficial for the prognosis prediction of LUAD patients, and was also expected to assist physicians in individualized treatment.

1. Introduction

Lung cancer is the primary contributor to cancer-related mortality on a global scale, with lung adenocarcinoma (LUAD) representing the predominant subtype [1,2]. The tumor-node-metastasis (TNM) staging system has the capability to predict the survival and risk stratification of patients [3], although its accuracy is limited [4]. Recent studies have shown that the assessment of the tumor immune microenvironment (TIME) in solid tumors is a beneficial complement to the TNM staging system [5]. The status of TIME is closely related to tumor progression, metastasis, and prognosis [6,7]. TIME is primarily affected by the interactions between tumor-infiltrating lymphocytes (TILs) and tumor cells. At present, many research fields [8–11] have demonstrated that the interaction between TILs and tumor cells is associated with the prognosis of patients. Therefore, it is anticipated that the interactions between TILs and tumor cells will be characterized in order to identify a novel biomarker for prognosis prediction and risk stratification in LUAD.

TILs have the potential to be valuable in predicting the effectiveness of the immune response and prognosis [12]. Many studies have concentrated on the complex spatial interplay between TILs and tumor cells, which could exert an influence on the occurrence, development, invasion, and metastasis of cancers [10,11,13]. In fact, the spatial distribution of cells is the prerequisite and important influencing factor for the interactions between them. Some studies [14,15] have indicated that the spatial distribution structure of cells is closely associated with tumor treatment response and patient prognosis. In particular, the distance between various types of cell subgroups is indicative of the prognostic effect of immunotherapy. However, these studies have mainly used multiplex immunofluorescence (mIF) or immunohistochemical (IHC) to characterize the relationship between TILs and tumor cells. These methods are

Table 1
Distributions of clinicopathologic factors of all three cohorts.

	D1	V1	V2	*p*-value
Age at surgery (median [IQR])	62.0 (55.0, 69.0)	62.0 (57.0, 66.5)	54.0 (48.0, 62.5)	<0.001[1]
<65 years	181 (65.6 %)	97 (69.8 %)	98 (85.2 %)	<0.001[2]
≥65 years	95 (34.4 %)	42 (30.2 %)	17 (14.8 %)	
Sex				0.083[2]
Male	141 (51.1 %)	86 (61.9 %)	58 (50.4 %)	
Female	135 (48.9 %)	53 (38.1 %)	57 (49.6 %)	
Smoking status				<0.001[2]
Never	208 (75.4 %)	65 (46.8 %)	74 (64.3 %)	
Former/current	68 (24.6 %)	74 (53.2 %)	41 (35.7 %)	
pT stage				<0.001[2]
T1	124 (44.9 %)	42 (30.2 %)	81 (70.4 %)	
T2	130 (47.1 %)	60 (43.2 %)	21 (18.3 %)	
T3	19 (6.9 %)	24 (17.3 %)	6 (5.2 %)	
T4	3 (1.1 %)	13 (9.3 %)	7 (6.1 %)	
pN stage				<0.001[2]
N0	218 (79.0 %)	80 (57.6 %)	83 (72.2 %)	
N1	18 (6.5 %)	19 (13.7 %)	13 (11.3 %)	
N2	40 (14.5 %)	40 (28.7 %)	19 (16.5 %)	
TNM stage				<0.001[2]
I	198 (71.7 %)	50 (36.0 %)	70 (60.9 %)	
II	32 (11.6 %)	33 (23.7 %)	17 (14.8 %)	
III	46 (16.7 %)	56 (40.3 %)	28 (24.3 %)	
Tumor site				<0.001[2]
Upper/Middle lobe	183 (66.3 %)	76 (54.7 %)	65 (56.5 %)	
Lower lobe	93 (33.7 %)	63 (45.3 %)	50 (43.5 %)	
Adjuvant chemotherapy				<0.001[2]
No	211 (76.4 %)	63 (45.3 %)	56 (48.7 %)	
Yes	65 (23.6 %)	76 (54.7 %)	59 (51.3 %)	
Follow-up (month, median [95 % CI])[a]	85.0 (82.4–88.6)	36.9 (36.4–43.0)	60.8 (57.8–65.1)	
No. of OS events	62 (22.5 %)	46 (33.1 %)	30 (26.1 %)	
No. of DFS events	116 (42.0 %)	67 (48.2 %)	48 (41.7 %)	

Notes: IQR stands for Interquartile Range, and CI stands for Confidence Interval. [1],[2] Since the data used in this paper is a subset of [22], so *p*-values determination is as same as this literature.

[a] Estimated by reverse K-M method.

costly and susceptible to variation due to the selection of antibodies, the experimental procedures, and the utilization of diverse automated immunohistochemical staining instruments.

Hematoxylin and Eosin (H&E) -stained images contain abundant TIME information, which is expected to be a beneficial supplement to mIF/IHC analysis. Conventionally, pathologists manually evaluate the TILs of patients on H&E slides [16]. However, manual quantification is labour-intensive and subjective, which often leads to interobserver inconsistency [17]. The advent of computer technology has spurred the advancement of automated analysis of whole-slide images (WSIs). Deep learning (DL) technology has the potential to extract image features automatically providing a valuable tool for quantitative analysis of WSIs [18], information mining [19], and prognostic analysis [20]. This technology has the potential to automate the quantification of TILs in WSIs. In addition, pathomic features (e.g. spatial arrangement features) provide additional information that may be useful for TILs detection, thus utilizing DL methods to analyze these features for prognosis is also a current research focus [17,21]. Nevertheless, recent workflows are still insufficient in exploring effective biomarkers of the spatial distance between cells for prognostic value based on H&E-stained WSIs. From this perspective, it is anticipated that a robust biomarker will be identified which quantifies the spatial distance between tumor cells and lymphocytes in patients with LUAD.

In this study, we proposed a simple but effective deep learning-driven workflow to segment and classify the tumor cells and TILs from H&E-stained WSIs. We then quantified the spatial distance between tumor cells and lymphocytes (referred to as DIST) based on the DL outputs. Finally, a prognosis analysis of this proposed biomarker was performed in patients with LUAD from three independent centers. Our hypothesis is that DIST is an independent prognostic factor for disease-free survival (DFS) and overall survival (OS), and that it could provide effective prognostic stratification of patients with resectable LUAD.

2. Method

2.1. Patients

Patients with LUAD were retrospectively enrolled from three medical institutions: GDPH (Discovery cohort 1, D1), SXCH (Validation cohort 1, V1), and YNCX (Validation cohort 2, V2). The clinicopathological information and H&E-stained WSIs of patients were collected. The clinicopathological information of patients is listed in Table 1. Due to the retrospective image analysis approach utilized in this study, the requirement for patient informed consent was waived. The specific inclusion and exclusion criteria are shown in Fig. 1. The endpoints of interest for this study are DFS and OS, which are defined as the time from surgery to the first recurrence or death and the time from surgery to death, respectively.

2.2. WSI digitalization and quality control

The H&E-stained glass sections of patients were scanned as WSIs using two different scanners, as shown in Fig. 2(a). The first scanner was a Leica Aperio-AT2 scanner (USA, 40× magnification, 0.25 μm/pixel) in D1 and V2, and the second was a NanoZoomer-NDP scanner (Japan, 40× magnification, 0.22 μm/pixel) in V1. Given the considerable heterogeneity in the characteristics of WSIs across different institutes, the WSIs used in this study underwent a meticulous review by two senior pathologists. This process involved the selection of the most representative slides and the removal of low-quality samples, such as those exhibiting blurring or out-of-focus characteristics.
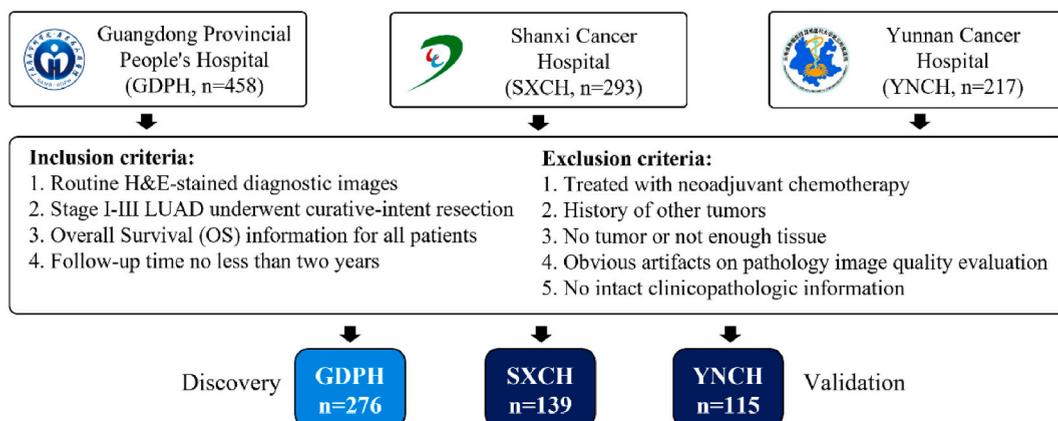

Fig. 1. Inclusion criteria and exclusion criteria. Discovery and validation cohort inclusion criteria and exclusion criteria in this study. Inclusion criteria: 1) Routine H&E-stained diagnostic images, 2) Stage I-III LUAD underwent curative-intent resection, 3) Overall Survival (OS) information for all patients, 4) Follow-up time no less than two years. Exclusion criteria: 1) Treated with neoadjuvant chemotherapy, 2) History of other tumors, 3) No tumor or not enough tissue, 4) Obvious artifacts on pathology image quality evaluation, 5) No intact clinicopathologic information.
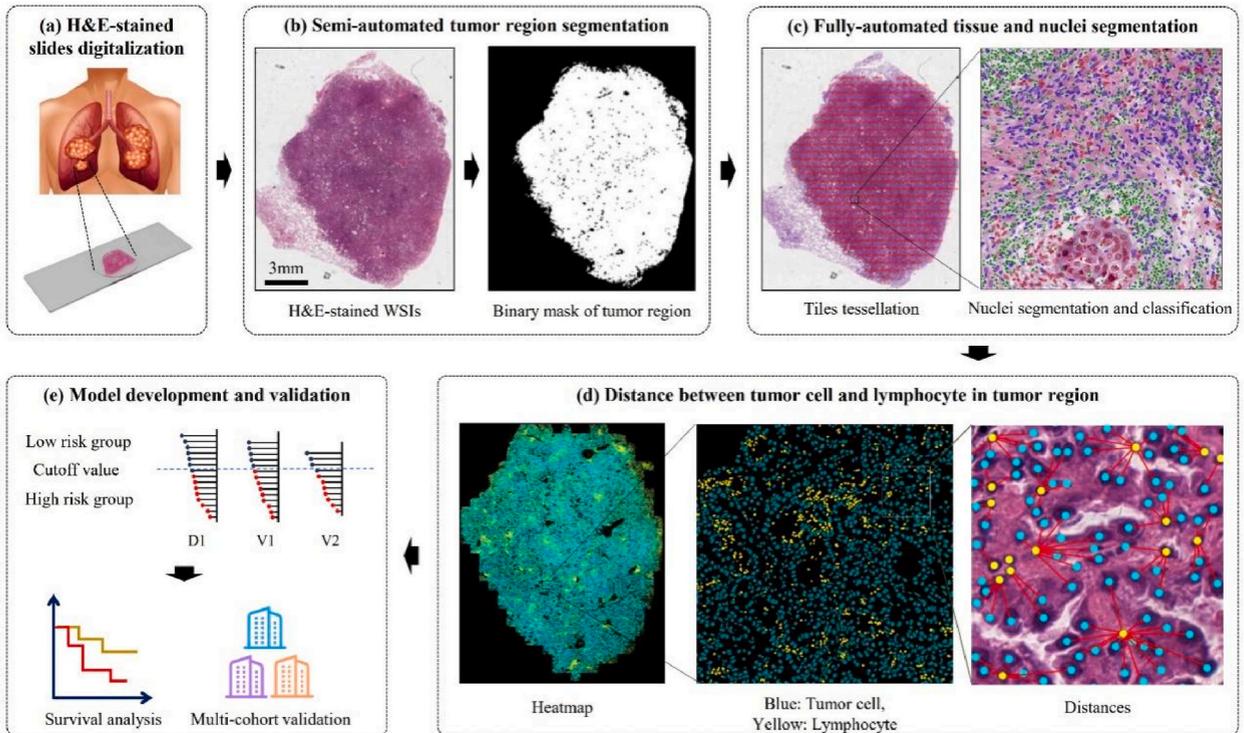
Fig. 2. The overall workflow of the study. (a) H&E-stained slides were digitized into whole slide images (WSIs). (b) The tumor region of WSI was segmented by a transfer learning framework, and segmentation masks were checked by two pathologists. (c) Fully automated nuclei segmentation and classification. HoVer-Net was deployed to segment and classify the cells into tumor cells, lymphocytes, and other cells in the tumor region of WSI at the maximum magnification. (d) Distance between the tumor cell and lymphocyte in the tumor region. In each WSI, the distance of the nearest lymphocyte to the tumor cell was calculated, and then we averaged the distance (DIST) of all tumor cells as the imaging biomarker. (e) Model development and validation. A discovery cohort was used to optimize DIST and the optimal value in the discovery cohort was selected as the cut-off to stratify patients into low and high DIST groups. Univariable and multivariable Cox regression analyses were performed to evaluate the association of DIST and clinical risk factors with DFS and OS. The other two cohorts were used to validate the performance of the model.

2.3. Tumor region segmentation

During the digitization of the WSI, a significant quantity of white background will be generated. These regions are of little value in the context of pathological image analysis and will result in a significant increase in the amount of calculation required. To remove the white background in the WSI and obtain the tumor region of interest, thesubsequent steps were implemented to conduct tumor region segmentation. Firstly, the OTSU operation was performed on WSI at low magnification to obtain the tissue mask. Secondly, the deep convolutional network ResNet50, which had been pre-trained on the Camelyon Challenge dataset [23], was employed to segment the WSI into the tumor lesion regions and other regions, as shown in Fig. S1. Subsequently, 67 WSIs previous annotated by pathologists from discovery cohort were selected to finetune the pre-trained model. Following finetuning, pathologists proceeded to manually correct any inaccuracies in the segmented tumor regions. The finetuning model was then deployed to the remaining WSIs in D1 and the other two cohorts for tumor region segmentation. Finally, the tumor region was segmented into patches (with a size of 2048×2048) with non-overlapping boundaries. For each patch, the model predicted the classification probability, thereby obtaining a probability result of the WSI. This result was binarized by the OTSU operation, and the final tumor region was determined as the largest connected area, as shown in Fig. 2(b).

2.4. Nuclei segmentation and classification

In this study, we utilized the Hover-Net [24], which had been pre-trained on the PanNuke dataset [25], for the purpose of segmenting and classifying nuclei within tumor regions. Four categories of nuclei were segmented and classified, including tumor cells, lymphocytes, stroma cells, and others. The entire process is shown in Fig. 2(c).

2.5. TILs density calculation

The quantification of the density of TILs can be formulated as Eqs. (1) and (2):

$$\text{DLE} = \frac{\text{Number of lymphocytes in tumor epithelium}}{\text{Area of tumor epithelium}} \tag{1}$$

$$\text{DLS} = \frac{\text{Number of lymphocytes in tumor} - \text{associated stroma}}{\text{Area of tumor} - \text{associated stroma}} \tag{2}$$

where DLE and DLS represented the density of TILs in the tumor epithelium and tumor-associated stroma, respectively. Subsequently, a base-2 logarithmic operation was applied to the TILs density for the purpose of enhanced analysis. The process above can be summarized as Eqs. (3) and (4):

$$\text{DLE} - \text{score} = \log_2(\text{DLE} + 1) \tag{3}$$

$$\text{DLS} - \text{score} = \log_2(\text{DLS} + 1) \tag{4}$$

2.6. Spatial distance quantification

The spatial distance (DIST) between tumor cells and lymphocytes was established through three steps. Firstly, the position co-ordinates (X, Y) of each patch in the WSI obtained in Section 2.3 were recorded, and the coordinate (x,y) of each cell center was also recorded separately. Then, the centroid coordinates of each cell in the WSI was converted to absolute coordinates according to the position coordinate of each patch in the WSI and the coordinate of the cell center in that patch. Secondly, the coordinates of the tumor cells and lymphocytes were spliced, with the absolute coordinates of each tumor cell defined as a node. The number of lymphocytes surrounding the tumor node was then calculated at a specific threshold (25 μm/pixel for D1 and V2, and 22 μm/pixel for V1), as shown in Fig. 2(d). Finally, we assigned a number to each tumor cell (1, 2, i, …, m) in each WSI, and calculated the spatial distance of the nearest lymphocyte to each tumor cell i named d_i using the 2-norm method. We then averaged the distances d_{avg} from all m tumor cells to the nearest lymphocyte as Eq. (5):

$$d_{avg} = \frac{1}{m} \sum_{i=1}^{m} d_i \tag{5}$$

The quantification of spatial distance between tumor cells and lymphocytes at the WSI level can be achieved using the method described above. To further analyze the effect of spatial distance (DIST) between tumor cells and lymphocytes, we sorted all DIST values for each patient in three cohorts (D1, V1, and V2). The optimal cut-off was then determined based on OS for D1, which divided patients into the low DIST group and the high DIST group.

2.7. Immune phenotypes determination

The immune phenotypes were grouped by immune-desert, immune-excluded, and inflamed. Amongst these, the immune-desert phenotype was linked to a poor prognosis, whereas the inflamed phenotype was correlated to a favorable prognosis [12,26]. To investigate the relationship between DIST and immune phenotypes, patients were divided into three cohorts into different immune phenotypes based on the density of TILs in the tumor epithelium and tumor-associated stroma, as described in Section 2.5. As in Ref. [26], samples with a lower tertile cut-off value for the DLE-score were classified as immune-desert. The remaining samples were divided into two groups, based on the median value of the ratio of the DLE-score and the DLS-score, which was used as a cut-off. The group above the cut-off was labelled as inflamed, and another was labelled as immune-excluded.

2.8. Statistical analysis

Continuous variables are depicted as the median (interquartile range). Discrete variables are expressed as counts (percentage). Comparative analysis across different centers is conducted using the Kruskal-Wallis rank-sum test for continuous variables and Pearson's Chi-square test or Fisher's exact test for discrete variables. The Cox proportional hazards regression model is used for survival analysis. The Kaplan–Meier (K-M) survival curves between the high-risk and low-risk groups are compared using the log-rank test. The factors that show statistical significance ($p < 0.05$) in the univariable Cox regression analysis will be selected as potential candidates for the multivariable Cox regression analysis. Harell's C-statistics (C-index) and the integrated area under the curve (iAUC) are used to evaluate the discriminant performance of the predictive model.

2.9. Implementation details

The deep learning-driven methodologies are implemented by PyTorch framework (Version 1.7.1; https://pytorch.org/) on an Ubuntu 18.04 server equipped with an 3090 GPU. The implementation of tumor segmentation is released at https://github.com/YuMeng-W/TumSeg-main. The experiment conditions of nuclei segmentation and classification are as same as the settings reported in literature [24]. Statistical analyses are conducted by R (Version 4.3.1; https://r-project.org/), and relevant R packages including survivor, survminer, timeROC, and Rmisc. All statistical tests adopt a two-tailed approach, wherein asignificance level of $p < 0.05$ is considered statistically significant.

3. Results

3.1. Baseline characteristics of patients

Following the meticulous application of all inclusion and exclusion criteria delineated in Fig. 1, the eligible patients were duly enrolled in each cohort. Table 1 outlined the comprehensive details pertaining to the clinicopathological characteristics of three cohorts. Statistically significant differences were observed among the three cohorts in terms of clinicopathological factors. (i.e., age at surgery, smoking status, pT stage, pN stage, TNM stage, tumor site, and adjuvant chemotherapy), with the exception of sex ($p = 0.083$).

3.2. Prognostic ability of DIST

We divided the patients with LUAD into three independent cohorts according to Section 2.1. The maximally selected rank statistics [27] were employed in the discovery cohort to determine the potential cut-off value of DIST. Consequently, a patient with a DIST value exceeding the cut-off value was categorized as belonging to the high DIST group, and vice versa. Kaplan-Meier (KM) curves
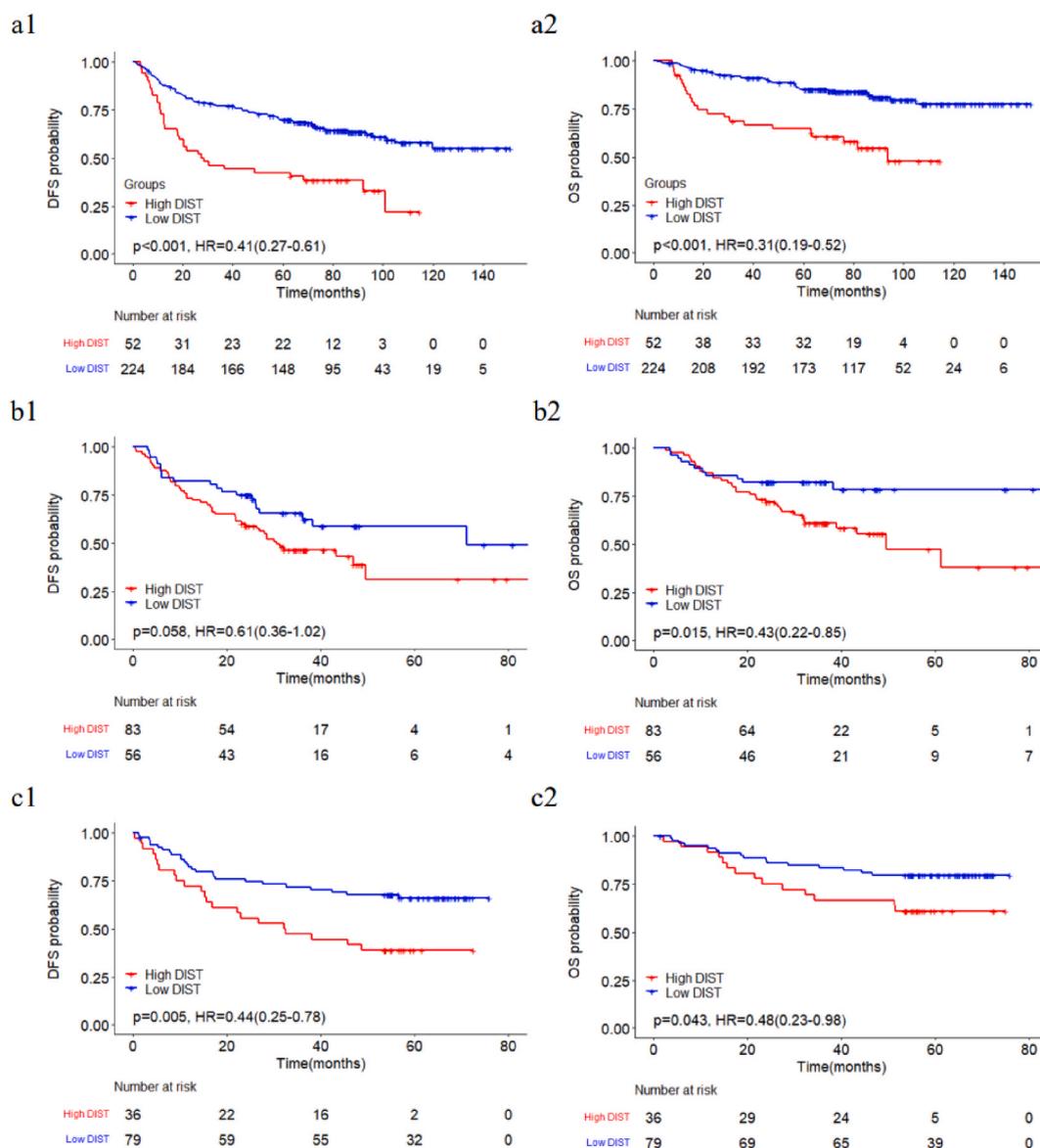

Fig. 3. Kaplan-Meier curves of patients in three cohorts in terms of DFS and OS. Kaplan-Meier curves of patients stratified by DIST in the discovery cohort D1 (a1) and the two validation cohorts V1 and V2 (b1 and c1) in terms of DFS. Kaplan-Meier curves of patients stratified by DIST in discovery cohort D1 (a2) and two validation cohorts V1 and V2 (b2 and c2) in terms of OS.

demonstrated that the low DIST group exhibited superior DFS than the high DIST group (D1, hazard ratio [HR], 0.41; 95 % confidence interval [CI], 0.27–0.61; p < 0.001; V1, HR, 0.61; 95 % CI, 0.36–1.02; p = 0.058; V2, HR, 0.44; 95 % CI, 0.25–0.78; p = 0.005; Fig. 3 (a1)–(c1)). In terms of OS, the low DIST group had better survival rate compared to the high DIST group (D1, HR, 0.31; 95 % CI, 0.19–0.52; p < 0.001; V1, HR, 0.43; 95 % CI, 0.22–0.85; p = 0.015; V2, HR, 0.48; 95 % CI, 0.23–0.98; p = 0.043; Fig. 3(a2)–(c2)). For the subgroup analysis, the data from three centers were combined. In terms of DFS, it was observed that patients belonging to the low DIST group exhibited superior DFS compared to those in the high DIST group. This was statistically significant in subgroup analysis for age at surgery, sex, smoking status, and TNM stage (p < 0.05). However, there was a marginal statistical significance in adjuvant chemotherapy (p = 0.32; Fig. S2). Similarly, patients belonging to the low DIST group had better OS than their counterparts in the high DIST group. This was statistically significant in the age at surgery, sex, smoking status, TNM stage, and received adjuvant chemotherapy subgroups (all p < 0.05; Fig. S3).

In terms of DFS, the experimental results of univariable and multivariable Cox regression analyses are shown in Table 2. In the discovery cohort, the univariable Cox regression analysis showed that factors (age at surgery, sex, smoking status, TNM stage, and DIST) exhibited statistical significance with p < 0.05. After adjustment for relevant clinical factors, the multivariable Cox regression analysis distinctly indicated that DIST emerged as an independent prognostic factor for DFS (D1, HR, 0.61; 95 % CI, 0.40–0.94; p = 0.027; V1, HR, 0.54; 95 % CI, 0.32–0.91; p = 0.022; V2, HR, 0.44; 95 % CI, 0.24–0.81; p = 0.009). The experimental results also showed that DIST emerged as an independent prognostic factor for OS (D1, HR, 0.42; 95 % CI, 0.24–0.73; p = 0.002; V1, HR, 0.36; 95 % CI, 0.18–0.72; p = 0.004; Table 3).

3.3. Prognostic model development and validation

The whole process of model development and validation is shown in Fig. 2(e). Considering age at surgery, TNM stage, and DIST as independent prognostic factors in the multivariable Cox regression analysis of the discovery cohort pertaining to DFS, we selected the above factors to develop a prediction model (named full model). We used the full model to compare and analyze with the TNM stage model, DIST model, TNM_DIST model (TNM stage and DIST factors jointly modeled), and Clinicopathologic model (age at surgery and TNM stage jointly modeled). A comprehensive summary encompassing the baseline hazard, coding, and model's regression coefficients was provided in Table S1 (DFS) and Table S2 (OS). In addition, the discriminant performance of each model was evaluated by the C-index and iAUC, and the correction performance of each model was evaluated by the Akaike Information Criterion (AIC). In Table 4 and Table S3, the full model had stronger discriminative and correction performance than the clinicopathological model in the discovery cohort and two validation cohorts. Specifically, the TNM_DIST model had stronger discriminative and correction performance compared to the TNM stage model subsequent to the inclusion of the DIST factor. Adding the DIST factor to the TNM stage model enhances the predictive capacity for DFS (likelihood ratio test, c^2 = 3.932, p = 0.047), and similarly, adding the DIST factor to

Table 2
Univariable and multivariable Cox regression analyses for DFS in three cohorts.

	Discovery cohort D1[a]		Validation cohort V1		Validation cohort V2	
	HR (95 % CI)	*p*-value	HR (95 % CI)	*p*-value	HR (95 % CI)	*p*-value
Univariable Analysis						
Age at surgery						
≥65 years vs. < 65 years	1.40 [0.97, 2.03]	0.076	0.77 [0.44, 1.34]	0.354	1.25 [0.58, 2.66]	0.571
Sex						
Female vs. Male	0.55 [0.38, 0.80]	0.002	0.56 [0.33, 0.95]	0.031	0.92 [0.52, 1.61]	0.762
Smoking status						
Former/Current vs. Never	1.79 [1.22, 2.64]	0.003	1.64 [1.00, 2.68]	0.049	0.66 [0.35, 1.23]	0.190
TNM stage						
III vs. I/II	4.91 [3.29, 7.33]	<0.001	2.28 [1.41, 3.70]	<0.001	2.97 [1.66, 5.32]	<0.001
Tumor site						
Upper/Middle vs. Lower	1.37 [0.92, 2.05]	0.121	1.08 [0.67, 1.75]	0.757	1.25 [0.70, 2.23]	0.449
DIST						
Low vs. High	0.41 [0.27, 0.61]	<0.001	0.61 [0.36, 1.02]	0.058	0.44 [0.25, 0.78]	0.005
Multivariable Analysis						
Age at surgery						
≥65 years vs. < 65 years	1.68 [1.15, 2.47]	0.007	0.81 [0.46, 1.42]	0.459	0.93 [0.41, 2.10]	0.864
Sex						
Female vs. Male	0.66 [0.42, 1.01]	0.058	0.63 [0.27, 1.47]	0.282	0.54 [0.25, 1.18]	0.123
Smoking status						
Former/Current vs. Never	1.45 [0.92, 2.27]	0.108	1.14 [0.51, 2.55]	0.741	0.36 [0.15, 0.83]	0.016
TNM stage						
III vs. I/II	4.60 [2.98, 7.10]	<0.001	2.31 [1.42, 3.77]	<0.001	2.51 [1.37, 4.59]	0.003
DIST						
Low vs. High	0.61 [0.40, 0.94]	0.027	0.54 [0.32, 0.91]	0.022	0.44 [0.24, 0.81]	0.009

Notes:
[a] D1 is the discovery cohort.

Table 3
Univariable and multivariable Cox regression analyses for OS in three cohorts.

	Discovery cohort D1[a]		Validation cohort V1		Validation cohort V2	
	HR (95 % CI)	*p*-value	HR (95 % CI)	*p*-value	HR (95 % CI)	*p*-value
Univariable Analysis						
Age at surgery						
≥65 years vs. < 65 years	2.04 [1.24, 3.36]	0.005	1.08 [0.57, 2.02]	0.814	1.39 [0.57, 3.41]	0.468
Sex						
Female vs. Male	0.49 [0.29, 0.83]	0.007	0.43 [0.22, 0.85]	0.015	0.95 [0.46, 1.95]	0.894
Smoking status						
Former/Current vs. Never	2.00 [1.19, 3.34]	0.008	2.04 [1.11, 3.75]	0.021	0.69 [0.32, 1.51]	0.351
TNM stage						
III vs. I/II	5.54 [3.30, 9.30]	<0.001	2.02 [1.13, 3.61]	0.018	3.48 [1.69, 7.14]	<0.001
Tumor site						
Upper/Middle vs. Lower	1.39 [0.80, 2.44]	0.243	1.38 [0.76, 2.51]	0.296	1.01 [0.49, 2.07]	0.986
DIST						
Low vs. High	0.31 [0.19–0.52]	<0.001	0.43 [0.22–0.85]	0.015	0.48 [0.23–0.98]	0.043
Multivariable Analysis						
Age at surgery						
≥65 years vs. < 65 years	2.76 [1.64, 4.64]	<0.001	1.28 [0.67, 2.44]	0.462	1.08 [0.42, 2.79]	0.866
Sex						
Female vs. Male	0.59 [0.32, 1.07]	0.083	0.57 [0.20, 1.61]	0.291	0.65 [0.24, 1.72]	0.382
Smoking status						
Former/Current vs. Never	1.43 [0.79, 2.60]	0.242	1.54 [0.60, 3.99]	0.369	0.46 [0.16, 1.29]	0.139
TNM stage						
III vs. I/II	5.16 [2.94, 9.06]	<0.001	1.94 [1.08, 3.48]	0.026	3.10 [1.49, 6.48]	0.003
DIST						
Low vs. High	0.42 [0.24, 0.73]	0.002	0.36 [0.18, 0.72]	0.004	0.52 [0.24, 1.11]	0.09

Notes:
[a] D1 is the discovery cohort.

Table 4
Performance of the full model and reference models in three cohorts (OS).

Cohort	Model	C-index (95 % CI)	iAUC	AIC
Discovery cohort (D1)	TNM stage	0.651 (0.592–0.710)	0.638	629.2
	DIST	0.614 (0.554–0.675)	0.606	647.2
	TNM_DIST	0.695 (0.631–0.759)	0.677	625.1
	Clinicopathologic	0.723 (0.658–0.788)	0.721	618.9
	Full	0.745 (0.682–0.809)	0.746	612.9
Validation cohort 1 (V1)	TNM stage	0.586 (0.511–0.661)	0.590	417.4
	DIST	0.570 (0.497–0.644)	0.598	416.3
	TNM_DIST	0.613 (0.524–0.702)	0.645	412.4
	Clinicopathologic	0.596 (0.512–0.680)	0.593	417.9
	Full	0.621 (0.527–0.716)	0.650	413.5
Validation cohort 2 (V2)	TNM stage	0.632 (0.547–0.717)	0.634	267.2
	DIST	0.585 (0.497–0.673)	0.586	273.9
	TNM_DIST	0.682 (0.593–0.771)	0.673	264.8
	Clinicopathologic	0.650 (0.556–0.744)	0.648	267.0
	Full	0.671 (0.581–0.762)	0.677	265.6

Clinicopathologic model = age at surgery + TNM stage; Full model = age at surgery + TNM stage + DIST.

the Clinicopathologic model also improved the prediction ability of DFS (likelihood ratio test, $c^2 = 4.752$, p = 0.029). Figs. S4 and S5 demonstrated the ROC curves (t = 60 months) and time-dependent AUC plotted over time. Compared to the reference model, the full model showed a higher AUC at most time points.

3.4. Correlation between DIST and immune phenotypes

To further evaluate the effects of DIST on the immune microenvironment, we combined three datasets to explore the correlation between DIST and immune phenotypes. The differences in DIST across three immune phenotypes (immune-desert, immune-excluded, and inflamed) in all discovery and validation cohorts were statistically significant by the Pearson Chi-square test (p < 0.001). The results showed that the low DIST group exhibited a higher prevalence of the inflamed phenotype (D1, 100 %; V1, 78 %; V2, 92.3 %), and the high DIST group displayed a higher proportion of the immune-desert phenotype (D1, 52.7 %; V1, 95 %; V2, 93.8 %), while the low DIST group (D1, 94.8 %; V1, 44 %; V2, 60 %) and the high DIST group was closed in the immune-excluded phenotype (V1, V2), as shown in Fig. 4.
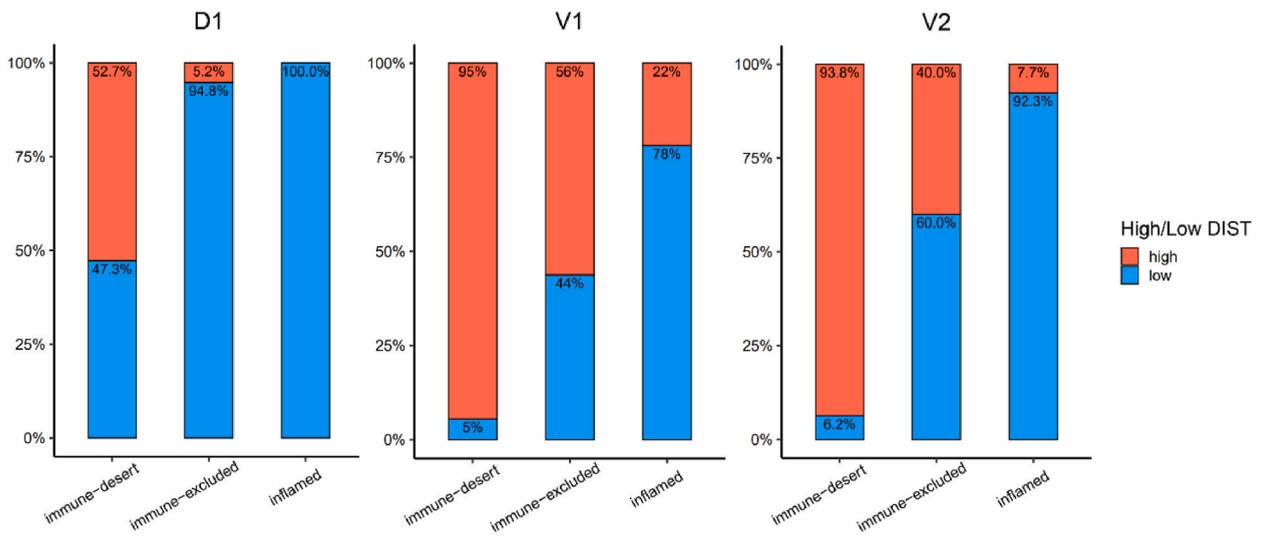
Fig. 4. Correlation between DIST and immune phenotypes in each cohort. The proportion of patients with different immune phenotypes correlated to different levels of DIST for each cohort are shown as histogram plots.
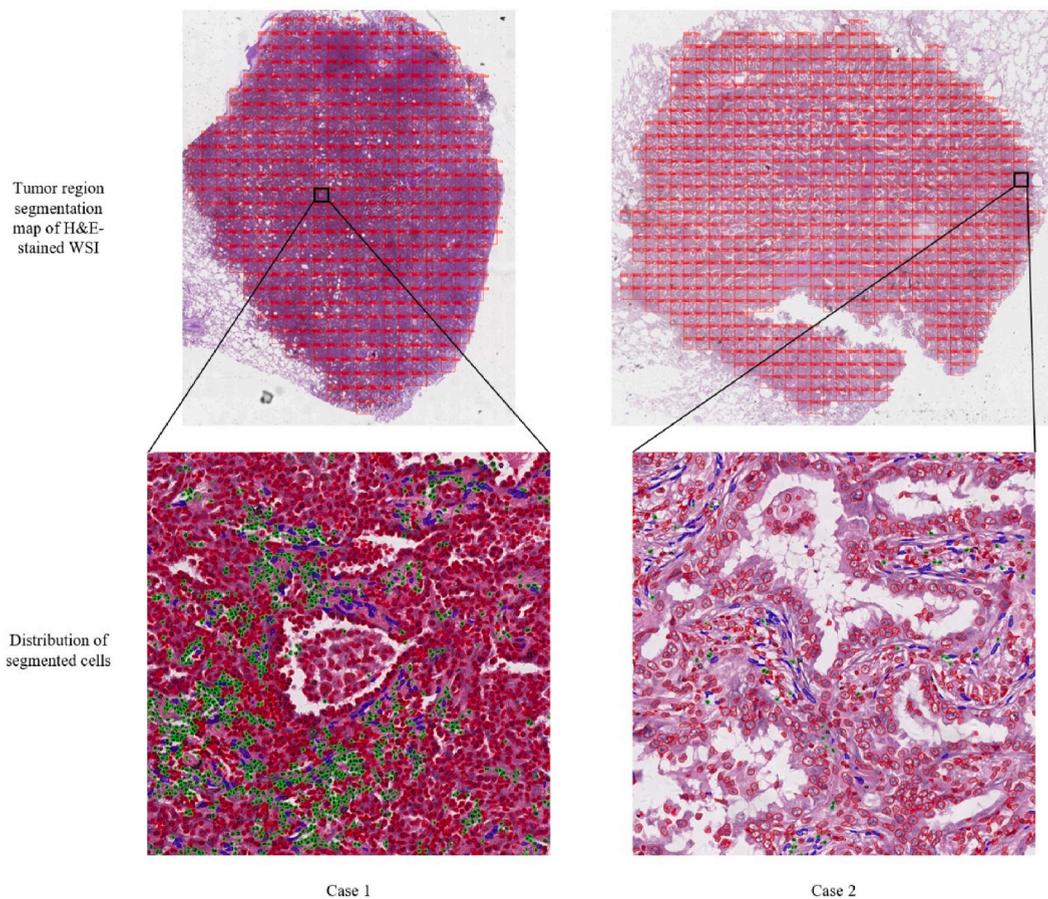

Fig. 5. Visualization of segmented cells in tumor region. The first row is the segmentation map of the tumor region, and the second row is the spatial distribution map of segmented cells in the tumor region. The red, green, blue, and black areas indicate tumor cells, lymphocytes, mesenchymal cells, and others, respectively.

3.5. Visualization of tumor cells and TILs in WSI

The central coordinates of tumor cells and lymphocytes are projected into pathological WSI, as illustrated in Fig. 5. The optimal cut-off value of DIST in the discovery cohort was 33.89 µm, and the DIST values of case 1 and case 2 were 19.87 and 43.69 µm, respectively. Therefore, case 1 and case 2 were classified as belonging to the low and high DIST groups, respectively. Follow-up data showed that case 1 had no recurrence at 130.2 months after the operation, while case 2 exhibited recurrence at 6.0 months after the operation. Both cases were classified as stage IA, yet exhibited markedly disparate DFS. The primary contributing factor was the degree of lymphocyte infiltration in the tumor region, which differed markedly between the two cases. Additionally, the distance between tumor cells and lymphocytes exhibited notable variability. As shown in Fig. 5, the lymphocytes of case 1 exhibited a higher density of infiltration within the tumor region, with a tendency to cluster at a greater distance from tumor cells.

4. Discussion

In this study, we proposed a framework for the automatic quantification of spatial distance between tumor cells and lymphocytes using H&E-stained WSIs. We calculated the average of the nearest lymphocyte distance of each tumor cell in the WSI (named DIST). The prognosis analysis found that compared with the high DIST group, the low DIST group had significantly better DFS and OS. The results also showed that DIST was an independent prognostic factor, and that it could be employed to provide effective prognostic stratification of patients with resectable LUAD.

Accurate tumor region segmentation and nuclei classification is the key prerequisite step for the spatial distance quantification between tumor cells and lymphocytes. Currently, the majority of studies evaluating the spatial interaction of TILs and tumor cells employed IHC- and mIF-stained pathological images to classify nuclei and characterize the relationship between them [10,28,29]. Galon et al. [28] characterized tumor-infiltrating immune cells in human colorectal cancer through gene expression profile and IHC image analysis. Lundgren et al. [10] used immune cell markers, such as CD4, CD8α, and FoxP3, to perform multiple immunofluorescence imaging on tumor tissue microarray (TMA) of 175 patients with periampullary adenocarcinoma. Our group calculated the density of CD3 and CD8 based on IHC pathological images and constructed the immune infiltration score named I-score, which could stratify patients with resectable NSCLC effectively and accurately [29]. However, these studies require a large amount of annotation data, and IHC- or mIF-stained pathological images are laborious and time-consuming to obtain. In addition, these images contain less TIME information compared to H&E-stained WSIs. In this study, we employed transfer learning to leverage publicly available labelled H&E-stained pathological images from other diseases for pre-training the model. This was then applied to lung cancer pathological images with only a few annotations, resulting in promising segmentation results. This approach significantly reduced the annotation workload of pathologists. Subsequently, the state-of-the-art model, Hover-Net [24], was employed in the process of accurate nuclei segmentation and classification using H&E-stained WSIs. The results demonstrated the significance of calculating the DIST from all tumor cells to the nearest lymphocyte in WSIs according to the coordinate of the cell, which would be conducive to downstream analysis.

The influence of spatial interactions between TILs and tumor cells on cancer prognosis has been extensively researched over many years. Most research have focused on the prognostic effect of in situ immune cell infiltration [30], the quantity of cells [7], pathological features [17,21], the density of TILs [22,31–33], and spatial patterns of TILs [34]. Federico et al. [30] conducted a comprehensive characterization of tissue slides from patients with locally advanced NSCLC using high-dimensional flow cytometry of TILs in conjunction with multiple immunofluorescence and gene sequencing. Research analysis indicated that increased B cell infiltration was related to a favorable prognosis in NSCLC patients. Additionally, literature [7] reported that the characteristics of the interaction between fibroblasts, TILs, and tumor cells could be employed for the classification of ER + HER2-breast cancer. This study indicated that the spatial mixing numbers of fibroblasts and TILs were related to good prognosis, but patients with high fibroblasts and low TILs had poor survival. Corredor et al. [17] predicted the likelihood of recurrence in NSCLC by capturing the spatial TIL features to co-localize the TILs and nuclei clusters across the tissue slide image. Verdicchio et al. [21] developed an automated pipeline to extract various hand-crafted pathomic features for the classification of TILs in breast cancer WSIs. These methods aimed to investigate the microscopic patterns in pathological data by extracting features from pathological analysis, but the process of selecting a large number of useful pathological features is time-consuming. Donnem et al. [31] conducted an assessment on the density of CD8+ TILs within the stroma, which was established it as an independent prognostic factor of NSCLC. Lin et al. [32] classified the tumors of LUAD into several subtypes based on TILs density and mRNA expression of PD-1/PD-L1, and demonstrate that their approach could better stratify patients of different clinical outcomes. However, the spatial distance and distribution between TILs and surrounding tumor cells is the premise of spatial interaction between them, which is not considered in the aforementioned papers. Barua et al. [35] used the G-cross spatial distance distribution method to analyze the abundance and spatial distribution of various cell types in TMA, along with their intercellular proximity. The results indicated that heightened infiltration of regulatory T cells into the central tumor region independently predicted a decrease in the overall survival rate of patients with NSCLC. Peng et al. [14] developed a framework to analyze the cell interaction networks in the tumor microenvironment, and revealed the spatial distance and complex interplays between immune and tumor cells. This study showed that spatial distance and compartmentalization between CD133+ cancer stem cells and CD8+ T cells were emphasized to play a pivotal role in immune response and carry out risk stratification on patients with LUAD. Although these researches achieving promising results, the number of extracted features is large and feature selection is complex, which is prone to over-fitting. Therefore, we proposed a simple and effective biomarker with simple calculation and good generalization ability to characterize the spatial distance between tumor cells and lymphocytes. It was demonstrated that the proposed biomarker DIST was an independent prognostic factor between DFS and OS in several external validation cohorts. By integrating the

DIST and baseline characteristics of patients, the full model demonstrated superior discriminative and correction performance in three independent cohorts when compared to models without DIST. This indicated that DIST not only enhances the prediction ability of DFS but also provides effective prognostic stratification of patients with resectable LUAD.

The immune phenotypes have contributed to the remodelling of the immunosuppressive microenvironment, the determination of prognosis, and the informing of novel cancer immunotherapy [36]. In general, the immune phenotypes include inflamed, immune-excluded, and immune-desert in NSCLC [26]. According to the literature [12], the immune-desert phenotype is correlated with unfavorable prognosis, whereas the inflamed phenotype is linked to a good prognosis. The DIST in this study may be considered to reflect the spatial distance between tumor cells and lymphocytes, thus to represent the ability of the WSI to predict prognosis. It was found that the low DIST group was strongly correlated with the inflamed phenotype, while the high DIST group was strongly correlated with the immune-desert phenotype, by exploring the correlation between DIST and immune phenotypes. Furthermore, we observed that a considerable proportion of patients in the immune-desert group were categorized as low DIST in the discovery cohort. This discrepancy arises because our proposed DIST metric averages the distance between tumor cells and lymphocytes across the entire tumor region, not only encompassing both tumor epithelium and tumor-associated stroma which are used to define the immune phenotypes. Thus, it was plausible that numerous lymphocytes may exist within the tumor epithelium. It can be reasonably assumed that these lymphocytes would influence the DIST calculation and lead to the observed results in the immune-desert group of Fig. 4. Overall, the experimental findings effectively illustrate that the proposed biomarker is beneficial in improving the risk stratification of patients with LUAD.

This study also has some limitations. Firstly, this study is retrospective, and the distribution of clinicopathological factors across the three centers is markedly disparate, presenting a significant challenge to the generalizability of the model. In addition, the production process of pathological slides and scanner parameters differs across each cohort, resulting in a considerable variation in pathological images and a high risk of batch effect. Secondly, due to the inherent limitations of the deep learning-based nuclei segmentation method, there is a certain degree of bias in the TILs density and the spatial distance calculations. Moreover, this study only considers the distance of lymphocytes in proximity to tumor cells, and more complex spatial relationships between tumor cells and lymphocytes will be further explored in the future. Besides, a single pathological WSI only reflects two-dimensional information at the section level, which is insufficient to reflect the three-dimensional spatial structure relationship between the cells. Finally, it is not possible to claim which type of lymphocyte proximity to tumor cells regards the positive prognostic impact. In order to investigate this relationship further, it is necessary to use specific biomarkers of lymphocyte subtypes.

5. Conclusions

In conclusion, we propose a simple but effective biomarker, DIST, which is calculated by measuring the spatial distance between tumor cells and lymphocytes. This biomarker was able to stratify patients with resectable LUAD into low-risk and high-risk groups in terms of DFS and OS, with significant differences. DIST was found to be an independent prognostic factor for DFS and OS, and the clinicopathological model integrated by DIST demonstrated an improvement in the stratification performance for patients with resectable LUAD. This finding was beneficial for the prognosis prediction of LUAD patients and was also expected to assist physicians in the administration of individualized treatment.

Ethics statement

The research has been approved by the Ethical Board Committee of Guangdong Provincial People's Hospital, and the approval numbers are KY-Z-2021-030-02, 202106, and KY2020139.

Data availability statement

Data will be made available on request.

CRediT authorship contribution statement

Xipeng Pan: Writing – original draft, Supervision, Software, Funding acquisition, Conceptualization. **Siyang Feng:** Writing – original draft, Software, Methodology. **Yumeng Wang:** Writing – review & editing, Software, Methodology. **Jiale Chen:** Writing – original draft, Software. **Huan Lin:** Data curation. **Zimin Wang:** Writing – review & editing. **Feihu Hou:** Methodology. **Cheng Lu:** Writing – review & editing. **Xin Chen:** Writing – review & editing. **Zhenbing Liu:** Writing – review & editing. **Zhenhui Li:** Funding acquisition. **Yanfen Cui:** Data curation. **Zaiyi Liu:** Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e30779.

References

[1] A.A. Thai, B.J. Solomon, L.V. Sequist, J.F. Gainor, R.S. Heist, Lung cancer, Lancet 398 (2021) 535–554, https://doi.org/10.1016/S0140-6736(21)00312-3.

[2] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (2021) 209–249, https://doi.org/10.3322/caac.21660.

[3] M.B. Amin, F.L. Greene, S.B. Edge, C.C. Compton, J.E. Gershenwald, R.K. Brookland, L. Meyer, D.M. Gress, D.R. Byrd, D.P. Winchester, The Eighth Edition AJCC Cancer Staging Manual: Continuing to Build a Bridge from a Population-Based to a More "Personalized" Approach to Cancer Staging: the Eighth Edition AJCC Cancer Staging Manual, CA: A Cancer J. Clin. 67 (2017) 93–99, https://doi.org/10.3322/caac.21388.

[4] K. Sugawara, H. Yamashita, M. Urabe, Y. Uemura, Y. Okumura, K. Yagi, S. Aikou, Y. Seto, Combining nutritional status with TNM stage: a physiological update on gastric cancer staging for improving prognostic accuracy in elderly patients, Int. J. Clin. Oncol. 27 (2022) 1849–1858, https://doi.org/10.1007/s10147-022-02250-5.

[5] M. Binnewies, E.W. Roberts, K. Kersten, V. Chan, D.F. Fearon, M. Merad, L.M. Coussens, D.I. Gabrilovich, S. Ostrand-Rosenberg, C.C. Hedrick, R.H. Vonderheide, M.J. Pittet, R.K. Jain, W. Zou, T.K. Howcroft, E.C. Woodhouse, R.A. Weinberg, M.F. Krummel, Understanding the tumor immune microenvironment (TIME) for effective therapy, Nat. Med. 24 (2018) 541–550, https://doi.org/10.1038/s41591-018-0014-x.

[6] T.N. Schumacher, D.S. Thommen, Tertiary lymphoid structures in cancer, Science 375 (n.d.) eabf9419. https://doi.org/10.1126/science.abf9419..

[7] N.K. Altorki, G.J. Markowitz, D. Gao, J.L. Port, A. Saxena, B. Stiles, T. McGraw, V. Mittal, The lung microenvironment: an important regulator of tumour growth and metastasis, Nat. Rev. Cancer 19 (2019) 9–31, https://doi.org/10.1038/s41568-018-0081-9.

[8] I. Nederlof, S. Hajizadeh, F. Sobhani, S.E.A. Raza, K. AbdulJabbar, R. Harkes, M.J. van de Vijver, R. Salgado, C. Desmedt, M. Kok, Y. Yuan, H.M. Horlings, Spatial interplay of lymphocytes and fibroblasts in estrogen receptor-positive HER2-negative breast cancer, Npj Breast Cancer 8 (2022) 56, https://doi.org/10.1038/s41523-022-00416-y.

[9] Ł. Zadka, M. Chabowski, D. Grybowski, A. Piotrowska, P. Dzięgiel, Interplay of stromal tumor-infiltrating lymphocytes, normal colonic mucosa, cancer-associated fibroblasts, clinicopathological data and the immunoregulatory molecules of patients diagnosed with colorectal cancer, Cancer Immunol. Immunother. 70 (2021) 2681–2700, https://doi.org/10.1007/s00262-021-02863-1.

[10] S. Lundgren, J. Elebro, M. Heby, B. Nodin, K. Leandersson, P. Micke, K. Jirström, A. Mezheyeuski, Quantitative, qualitative and spatial analysis of lymphocyte infiltration in periampullary and pancreatic adenocarcinoma, Int. J. Cancer 146 (2020) 3461–3473, https://doi.org/10.1002/ijc.32945.

[11] X. Wang, C. Barrera, K. Bera, V.S. Viswanathan, S. Azarianpour-Esfahani, C. Koyuncu, P. Velu, M.D. Feldman, M. Yang, P. Fu, others, Spatial interplay patterns of cancer nuclei and tumor-infiltrating lymphocytes (TILs) predict clinical benefit for immune checkpoint inhibitors, Sci. Adv. 8 (2022) eabn3966.

[12] S. Park, C.-Y. Ock, H. Kim, S. Pereira, S. Park, M. Ma, S. Choi, S. Kim, S. Shin, B.J. Aum, K. Paeng, D. Yoo, H. Cha, S. Park, K.J. Suh, H.A. Jung, S.H. Kim, Y. J. Kim, J.-M. Sun, J.-H. Chung, J.S. Ahn, M.-J. Ahn, J.S. Lee, K. Park, S.Y. Song, Y.-J. Bang, Y.-L. Choi, T.S. Mok, S.-H. Lee, Artificial intelligence–powered spatial analysis of tumor-infiltrating lymphocytes as complementary biomarker for immune checkpoint inhibition in non–small-cell lung cancer, J. Clin. Orthod. 40 (2022) 1916–1928, https://doi.org/10.1200/JCO.21.02010.

[13] Z. Feng, H. Lin, Z. Liu, L. Yan, Y. Wang, B. Li, E. Liu, C. Han, Z. Shi, C. Lu, Z. Liu, C. Pang, Z. Li, Y. Cui, X. Pan, X. Chen, Artificial intelligence-quantified tumour-lymphocyte spatial interaction predicts disease-free survival in resected lung adenocarcinoma: a graph-based, multicentre study, Comput. Methods Progr. Biomed. 238 (2023) 107617, https://doi.org/10.1016/j.cmpb.2023.107617.

[14] H. Peng, X. Wu, S. Liu, M. He, C. Xie, R. Zhong, J. Liu, C. Tang, C. Li, S. Xiong, H. Zheng, J. He, X. Lu, W. Liang, Multiplex immunofluorescence and single-cell transcriptomic profiling reveal the spatial cell interaction networks in the non-small cell lung cancer microenvironment, Clin. Transl. Med. 13 (2023) e1155, https://doi.org/10.1002/ctm2.1155.

[15] F. Zhou, G. Shayan, S. Sun, X. Huang, X. Chen, K. Wang, Y. Qu, R. Wu, Y. Zhang, Q. Liu, J. Zhang, J. Luo, X. Shi, Y. Liu, B. Liang, Y.-X. Li, J. Wang, J. Yi, Spatial architecture of regulatory T-cells correlates with disease progression in patients with nasopharyngeal cancer, Front. Immunol. 13 (2022), https://doi.org/10.3389/fimmu.2022.1015283.

[16] M. Rakaee, T.K. Kilvaer, S.M. Dalen, E. Richardsen, E.-E. Paulsen, S.M. Hald, S. Al-Saad, S. Andersen, T. Donnem, R.M. Bremnes, L.-T. Busund, Evaluation of tumor-infiltrating lymphocytes using routine H&E slides predicts patient survival in resected non–small cell lung cancer, Hum. Pathol. 79 (2018) 188–198, https://doi.org/10.1016/j.humpath.2018.05.017.

[17] G. Corredor, X. Wang, Y. Zhou, C. Lu, P. Fu, K. Syrigos, D.L. Rimm, M. Yang, E. Romero, K.A. Schalper, V. Velcheti, A. Madabhushi, Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non–small cell lung cancer, Clin. Cancer Res. 25 (2019) 1526–1534, https://doi.org/10.1158/1078-0432.CCR-18-2013.

[18] C. Han, H. Yao, B. Zhao, Z. Li, Z. Shi, L. Wu, X. Chen, J. Qu, K. Zhao, R. Lan, C. Liang, X. Pan, Z. Liu, Meta multi-task nuclei segmentation with fewer training samples, Med. Image Anal. 80 (2022) 102481, https://doi.org/10.1016/j.media.2022.102481.

[19] X. Pan, J. Cheng, F. Hou, R. Lan, C. Lu, L. Li, Z. Feng, H. Wang, C. Liang, Z. Liu, X. Chen, C. Han, Z. Liu, SMILE: cost-sensitive multi-task learning for nuclear segmentation and classification with imbalanced annotations, Med. Image Anal. 88 (2023) 102867, https://doi.org/10.1016/j.media.2023.102867.

[20] C.F. Koyuncu, C. Lu, K. Bera, Z. Zhang, J. Xu, P. Toro, G. Corredor, D. Chute, P. Fu, W.L. Thorstad, F. Faraji, J.A. Bishop, M. Mehrad, P.D. Castro, A.G. Sikora, L. D.R. Thompson, R.D. Chernock, K.A.L. Kuhs, J. Luo, V. Sandulache, D.J. Adelstein, S. Koyfman, J.S.L. Jr, A. Madabhushi, Computerized tumor multinucleation index (MuNI) is prognostic in p16+ oropharyngeal carcinoma, J. Clin. Investig. 131 (2021), https://doi.org/10.1172/JCI145488.

[21] M. Verdicchio, V. Brancato, C. Cavaliere, F. Isgrò, M. Salvatore, M. Aiello, A pathomic approach for tumor-infiltrating lymphocytes classification on breast cancer digital pathology images, Heliyon 9 (2023) 1.

[22] X. Pan, H. Lin, C. Han, Z. Feng, Y. Wang, J. Lin, B. Qiu, L. Yan, B. Li, Z. Xu, Z. Wang, K. Zhao, Z. Liu, C. Liang, X. Chen, Z. Li, Y. Cui, C. Lu, Z. Liu, Computerized tumor-infiltrating lymphocytes density score predicts survival of patients with resectable lung adenocarcinoma, iScience 25 (2022) 105605, https://doi.org/10.1016/j.isci.2022.105605.

[23] G. Litjens, P. Bandi, B. Ehteshami Bejnordi, O. Geessink, M. Balkenhol, P. Bult, A. Halilovic, M. Hermsen, R. van de Loo, R. Vogels, Q.F. Manson, N. Stathonikos, A. Baidoshvili, P. van Diest, C. Wauters, M. van Dijk, J. van der Laak, 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset, GigaScience 7 (2018) giy065, https://doi.org/10.1093/gigascience/giy065.

[24] S. Graham, Q.D. Vu, S.E.A. Raza, A. Azam, Y.W. Tsang, J.T. Kwak, N. Rajpoot, Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images, Med. Image Anal. 58 (2019) 101563, https://doi.org/10.1016/j.media.2019.101563.

[25] J. Gamper, N. Alemi Koohbanani, K. Benet, A. Khuram, N. Rajpoot, Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification, in: Digital Pathology: 15th European Congress, ECDP 2019, Warwick, UK, April 10–13, 2019, Proceedings 15, Springer, 2019, pp. 11–19.

[26] J. Park, H.-G. Cho, J. Park, G. Lee, H.S. Kim, K. Paeng, S. Song, G. Park, C.-Y. Ock, Y.K. Chae, Artificial intelligence–powered Hematoxylin and Eosin analyzer reveals distinct immunologic and mutational profiles among immune phenotypes in non–small-cell lung cancer, Am. J. Pathol. 192 (2022) 701–711, https://doi.org/10.1016/j.ajpath.2022.01.006.

[27] T. Hothorn, B. Lausen, On the exact distribution of maximally selected rank statistics, Comput. Stat. Data Anal. 43 (2003) 121–137, https://doi.org/10.1016/S0167-9473(02)00225-6.

[28] J. Galon, A. Costes, F. Sanchez-Cabo, A. Kirilovsky, B. Mlecnik, C. Lagorce-Pagès, M. Tosolini, M. Camus, A. Berger, P. Wind, F. Zinzindohoué, P. Bruneval, P.-H. Cugnenc, Z. Trajanoski, W.-H. Fridman, F. Pagès, Type, density, and location of immune cells within human colorectal tumors predict clinical outcome, Science 313 (2006) 1960–1964, https://doi.org/10.1126/science.1129139.

[29] H. Lin, X. Pan, Z. Feng, L. Yan, J. Hua, Y. Liang, C. Han, Z. Xu, Y. Wang, L. Wu, Y. Cui, X. Huang, Z. Shi, X. Chen, X. Chen, Q. Zhang, C. Liang, K. Zhao, Z. Li, Z. Liu, Automated whole-slide images assessment of immune infiltration in resected non-small-cell lung cancer: towards better risk-stratification, J. Transl. Med. 20 (2022) 261, https://doi.org/10.1186/s12967-022-03458-9.

[30] L. Federico, D.J. McGrail, S.-E. Bentebibel, C. Haymaker, A. Ravelli, M.-A. Forget, T. Karpinets, P. Jiang, A. Reuben, M.V. Negrao, J. Li, R. Khairullah, J. Zhang, A. Weissferdt, A.A. Vaporciyan, M.B. Antonoff, G. Walsh, S.-Y. Lin, A. Futreal, I. Wistuba, J. Roth, L.A. Byers, P.-O. Gaudreau, N. Uraoka, A.F. Cruz, H. Dejima, R.N. Lazcano, L.M. Solis, E.R. Parra, J.J. Lee, S. Swisher, T. Cascone, J.V. Heymach, J. Zhang, B. Sepesi, D.L. Gibbons, C. Bernatchez, Distinct tumor-infiltrating lymphocyte landscapes are associated with clinical outcomes in localized non-small-cell lung cancer, Ann. Oncol. 33 (2022) 42–56, https://doi.org/10.1016/j.annonc.2021.09.021.

[31] T. Donnem, S.M. Hald, E.-E. Paulsen, E. Richardsen, S. Al-Saad, T.K. Kilvaer, O.T. Brustugun, A. Helland, M. Lund-Iversen, M. Poehl, K.E. Olsen, H.J. Ditzel, O. Hansen, K. Al-Shibli, Y. Kiselev, T.M. Sandanger, S. Andersen, F. Pezzella, R.M. Bremnes, L.-T. Busund, Stromal CD8+ T-cell density—a promising supplement to TNM staging in non–small cell lung cancer, Clin. Cancer Res. 21 (2015) 2635–2643, https://doi.org/10.1158/1078-0432.CCR-14-1905.

[32] Z. Lin, J. Gu, X. Cui, L. Huang, S. Li, J. Feng, B. Liu, Y. Zhou, Deciphering microenvironment of NSCLC based on CD8+ TIL density and PD-1/PD-L1 expression, J. Cancer 10 (2019) 211.

[33] Y. Wang, H. Lin, N. Yao, X. Chen, B. Qiu, Y. Cui, Y. Liu, B. Li, C. Han, Z. Li, W. Zhao, Z. Wang, X. Pan, C. Lu, J. Liu, Z. Liu, Z. Liu, Computerized tertiary lymphoid structures density on H&E-images is a prognostic biomarker in resectable lung adenocarcinoma, iScience 26 (2023) 107635, https://doi.org/10.1016/j.isci.2023.107635.

[34] D.J. Fassler, L.A. Torre-Healy, R. Gupta, A.M. Hamilton, S. Kobayashi, S.C. Van Alsten, Y. Zhang, T. Kurc, R.A. Moffitt, M.A. Troester, others, Spatial characterization of tumor-infiltrating lymphocytes and breast cancer progression, Cancers 14 (2022) 2148.

[35] S. Barua, P. Fang, A. Sharma, J. Fujimoto, I. Wistuba, A.U.K. Rao, S.H. Lin, Spatial interaction of tumor cells and regulatory T cells correlates with survival in non-small cell lung cancer, Lung Cancer 117 (2018) 73–79, https://doi.org/10.1016/j.lungcan.2018.01.022.

[36] B. Zhao, Y. Wang, Y. Wang, C. Dai, Y. Wang, W. Ma, Investigation of genetic determinants of glioma immune phenotype by integrative immunogenomic scale analysis, Front. Immunol. 12 (2021), https://doi.org/10.3389/fimmu.2021.557994.