

Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images

Author names

Panwen Tian^{1*}, Bingxi He^{2,3*}, Di Dong^{3,4,5*}, Wei Mu^{6*}, Kunqin Liu⁷, Li Liu⁸, Hao Zeng⁹, Yujie Liu⁹, Lili Jiang¹⁰, Ping Zhou¹⁰, Zhipei Huang^{2†}, Weimin Li^{9†}, Jie Tian^{3,5,6,11,12†}.

Author affiliations

1 Department of Respiratory and Critical Care Medicine, Lung Cancer Treatment Centre, West China Hospital, West China Hospital, Sichuan University, Sichuan, China

2 School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China

3 CAS Key Laboratory of Molecular Imaging, Beijing Key Laboratory of Molecular Imaging, The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China

4 School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China

5 Zhuhai Precision Medical Center, Zhuhai People's Hospital (affiliated with Jinan University), Zhuhai, China

6 Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, China

7 Department of clinical medicine, Sichuan vocational college of health and rehabilitation, Zigong, Sichuan, China

8 Department of Respiratory Medicine, West China-Guang'an Hospital, Sichuan University, Sichuan, China

9 Department of Respiratory and Critical Care Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan, China

10 Department of Pathology, West China Hospital, Sichuan University, Chengdu, Sichuan, China

11 Key Laboratory of Big Data-Based Precision Medicine (Beihang University), Ministry of Industry and Information Technology, Beijing, China

12 Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi'an, Shaanxi, China

*** Panwen Tian, Bingxi He, Di Dong, and Wei Mu contributed equally to this work.**

†Corresponding author

Zhipei Huang, PhD

School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China.

Tel: +86 10 69671882

Email: zhphuang@ucas.ac.cn

Weimin Li, MD, PhD

Department of Respiratory and Critical Care Medicine, West China Hospital, Sichuan University. No.37 GuoXue Alley, Chengdu, Sichuan Province 610041, China

Tel: +86 28 85423998, **Fax:** +86 28 85582944

E-mail: weimin003@163.com

Jie Tian, PhD

The CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, No. 95 Zhongguancun East Road, Hai Dian District, Beijing, 100190, China.

Tel: +86 10 82618465, **Fax:** +86 10 62527995

Email: jie.tian@ia.ac.cn

Disclosure: The authors declare no conflict of interest.

Funding

This work was supported by the National Natural Science Foundation of China (91859203, 81871890, 82022036, 91959130, 81971776, 81771924, 81930053, 81227901, 81671851, 81527805), the National Key R&D Program of China (2017YFC0910004, 2017YFC1308700, 2017YFA0205200, 2017YFC1309100, 2017YFA0700401, 2017ZX10103004-012), the Beijing Natural Science Foundation (L182061), Strategic Priority Research Program of Chinese Academy of Sciences (XDB 38040200), the Instrument Developing Project of the Chinese Academy of Sciences (YZ201502), the Project of High-Level Talents Team Introduction in Zhuhai City (Zhuhai HLHPTP201703), the Youth Innovation Promotion Association CAS (2017175), the Bureau of International Cooperation of Chinese Academy of Sciences (173211KYSB20160053). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images

Abstract

Purpose:

This study aimed to use computed tomography (CT) images to assess PD-L1 expression in nonsmall cell lung cancer (NSCLC) and predict response to immunotherapy.

Methods:

We retrospectively analyzed a PD-L1 expression dataset that consisted of 939 consecutive stage IIIB-IV NSCLC patients with pretreatment CT images. A deep convolutional neural network was trained and optimized with CT images from the training cohort (n=750) and validation cohort (n=96) to obtain a PD-L1 expression signature (PDL1ES), which was evaluated using the test cohort (n=93). Finally, a separate immunotherapy cohort (n=94) was used to assess the prognostic value of PDL1ES with respect to clinical outcome.

Results:

PDL1ES was able to predict high PD-L1 expression ($\text{PD-L1} \geq 50\%$) with areas under the receiver operating characteristic curve (AUC) of 0.78 (95% confidence interval (CI): 0.75~0.80), 0.71 (95% CI: 0.60~0.81), and 0.76 (95% CI: 0.67~0.85) in the training, validation, and test cohorts, respectively. In patients treated with anti-PD-1 antibody, low PDL1ES was associated with improved progression-free survival (PFS) (median PFS 363 days in low score group vs 183 days in high score group; hazard ratio [HR]: 2.57, 95% CI: 1.22~5.44; $P = 0.010$). Additionally, when PDL1ES was combined with a clinical model that was trained using age, sex, smoking history and family history of malignancy, the response to immunotherapy ($P < 0.001$) could be better predicted compared to either PDL1ES or the clinical model alone.

Conclusions:

The deep learning model provides a noninvasive method to predict high PD-L1 expression of NSCLC and to infer clinical outcomes in response to immunotherapy. Additionally, this deep learning model combined with clinical models demonstrated improved stratification capabilities.

Key words: PD-L1 expression, deep learning, computed tomography, immunotherapy, nonsmall cell lung

cancer.

Introduction

Lung cancer remains the leading cause of cancer mortality worldwide [1]. Nonsmall cell lung cancer (NSCLC) represents 80-85% of primary lung malignancies [2]. The treatment strategy for NSCLC has rapidly evolved with the introduction of immune checkpoints inhibitors (ICIs) targeting PD-1 or PD-L1, but durable disease response remains limited to a subset of patients. Given the overall response rates to ICI treatment, ranging from 14% to 20% in unselected patients [3],[4], it is essential to identify predictive biomarkers for the selection of patients who are more likely to respond to immunotherapy.

PD-L1, which is expressed by cells in the tumor microenvironment, engages PD-1 on T cells and triggers inhibitory signaling of the T cell receptor, reducing T-cell killing capacity and blocking effector functions [5]. PD-L1 expression on tumor cells is induced by constitutive oncogenic signaling [6]. Alternatively, it can be induced in response to immune-stimulating cytokines, such as interferons, that are produced by an active antitumor immune response [7],[8]. Currently, PD-L1 expression on tumor cells assessed by immunohistochemistry is the only approved diagnostic biomarker for immunotherapy in patients with NSCLC and has been convincingly demonstrated to be associated with the efficacy of ICIs in NSCLC [9],[10]. In the first-line setting, pembrolizumab monotherapy demonstrated improved progression-free survival (PFS) and overall survival (OS) benefit compared to chemotherapy in NSCLC patients with PD-L1 expression on $\geq 50\%$ of tumor cells (KEYNOTE 024) [9],[11]. According to a recent study, though PFS and OS were significantly longer in the pembrolizumab-treated group than in the chemotherapy-treated group in the full NSCLC cohort with PD-L1 $\geq 1\%$, a clear PFS and OS benefit was only observed in the subgroup of patients with PD-L1 $\geq 50\%$, and higher PD-L1 expression corresponded to greater benefit (KEYNOTE 042) [12]. The FDA approved immunohistochemistry assay for PD-L1 expression, utilizing a cutoff of 50% tumor proportion score (TPS) for first-line treatment with pembrolizumab. Pembrolizumab monotherapy is preferred for patients with stage IV NSCLC and PD-L1 levels of 50% or more who are negative for *EGFR* mutations and *ALK* fusions. However, expression levels of PD-L1 are intratumorally heterogeneous and dynamic by immunohistochemistry analysis with different antibodies and platforms, as well as multiple scoring criteria, complicating interpretation of the results [13]-[15]. Therefore, a noninvasive and whole-tumor-based biomarker is urgently needed.

Radiological images are routinely available in clinical practice. Unlike traditional biopsy-based assays that represent only part of the tumor, images reflect information on the entire tumor burden in a non-invasive manner and avoid the effects of tumor heterogeneity [16],[17]. Radiomics is the science of quantifying patterns of tumor phenotypes on radiographic images in a high throughput manner and analyzing them with bioinformatics tools

to build clinically relevant models that assess tumor and microenvironment heterogeneity [19]. Indeed, radiomics-based biomarkers have shown success in predicting response to different treatments in different tumor types [21]-[26].

In this study, we aimed to develop a radiomic signature to predict PD-L1 TPS $\geq 50\%$ and to evaluate its potential ability to predict clinical outcomes in anti-PD-1 immunotherapy-treated metastatic NSCLC patients.

Patients and Methods

Patients

This study was granted ethics approval by the institutional review board of the West China Hospital, Sichuan University (Approval number: 2019-612) and was performed in accordance with ethical standards of the 1964 Declaration of Helsinki and its later amendments. Medical records were retrieved to collect clinical data (age, gender, smoking status, and family history), treatment regime, and survival outcome. Informed consent was waived due to the retrospective nature of this study.

PD-L1 expression dataset

A consecutive cohort of 2094 patients with stage IV NSCLC who underwent PD-L1 staining based on histological specimens from January 2016 to December 2018 at the West China Hospital of Sichuan University were retrospectively screened. Eight hundred twenty-four patients were excluded due to low-quality CT images. In this study, low-quality CT was defined as CT reconstructed with 30f reconstruction kernel. All included CTs were reconstructed with 60f kernel. Three hundred thirty-one patients were excluded due to insufficient clinical or pathological information. Finally, 939 patients were included in the analysis. Immunohistochemical (IHC) assays were performed using SP142 antibody on the Ventana Benchmark platform. Slides were scanned and independently scored by two pathologists who estimated the percentage of PD-L1 protein in both tumor cells and tumor infiltrating immune cells.

Immunotherapy dataset

Patients who met the following inclusion criteria were included in this dataset: 1) histologically confirmed primary NSCLC; 2) stage IV; 3) receipt of pembrolizumab monotherapy (200 mg every 3 weeks at the approved dose) or pembrolizumab combined with chemotherapy as the first-line treatment; and 4) pretreatment CT and CT data during follow-up were available. Patients were excluded if they met one or more of the following

criteria: 1) clinical data, including age, sex, smoking status and family history of tumor were missing; 2) other treatments, such as radiotherapy, surgery or herbs, were adopted during immunotherapy treatment; 3) low-quality CT images; or 4) lost to follow-up before disease progression. All enrolled patients submitted to CT scan two weeks before immunotherapy. The follow-up interval was 8–10 weeks, and routine laboratory tests and CT scans were performed during this time. Finally, 94 patients admitted between January 2017 to December 2018 from West China Hospital were included in the cohort. PFS was obtained using iRECIST criteria, which is the response criteria used in trials testing immunotherapeutic [27].

Data preprocessing

This study consisted of two datasets, the PD-L1 expression dataset and the immunotherapy dataset. Both datasets contained 4 clinical characteristics (age, gender, smoking status, and family history), as well as CT images. The PD-L1 expression dataset was used to train the PD-L1 expression signature (PDL1ES) program to reflect whether PD-L1 TPS is higher than 50%, while the immunotherapy dataset was used to evaluate the PFS utilizing the PDL1ES program. In this experiment, we obtained the training, validation and test cohorts from the PD-L1 expression dataset by stratified and random sampling of patients at a ratio of 8:1:1. Details of the datasets and experimental flow chart are shown in Supplementary Figure 1.

To make full use of the enrolled data, the k-nearest neighbor (KNN) algorithm was first performed to impute missing clinical values. In detail, the Euclidean distance between patients was calculated using patients' nonmissing variables, and missing values were filled by the weighted average of the closest K patients' value ($k = 3$ in this study). Then, all clinical characteristics were normalized to make the model easier to learn. All CT images were resampled to the same resolution of $1 \times 1 \times 1 \text{ mm}^3$.

Delineation of the region of interest (ROI)

The region of interest (ROI) of the primary tumor lesion was delineated by an experienced respiratory medicine specialist on the workstation. Next, we performed rectangular clipping of the ROI area amplifying to 5 pixels around to avoid the bias of ROI area delineation. Additionally, the upper and lower slices of the ROI slice were also cut and used as training sample to amplify the amount of available data. Furthermore, extracting the upper and lower slices at the same time also mitigates the error introduced by the radiologists' selection of slices. Finally, to unify the input size, we resized all ROI patches into 112×112 pixels. To more fully describe the tumor, we extracted radiomic features of ROIs to improve the accuracy of the model.

Establishment and evaluation of PD-L1 expression signature

We built the PDL1ES using a deep learning approach. The network (Figure 1) consisted of three parts, a deep learning feature extraction module based on the densenet121, a handcrafted conventional radiomic feature extraction module, and a classifier module based on the fully connected classification layer.

The input of the deep learning feature extraction module primarily included the slice of the CT tumor region. Before inputting the radiomic features, we employed the Mann Whitney U test to reduce the dimensionality of the radiomic features based on the training cohort. For clinical characteristics, we standardized them with z-score and concatenate them with radiomic features and deep learning features as the input of the fully connected layer. The output of the whole model represented the high expression probability and was regarded as PDL1ES. During training of the model, cross-entropy loss function and Adam optimizer were used, and the learning rates were set to 1e-4, 1e-5 and 1e-6. We dropped the learning rate by 20 epochs per training. The best standard for training in each step was the best result of the validation cohort.

The predictive value of the PDL1ES was evaluated with area under the receiver operating characteristic curve (AUC) and the decision curve. Meanwhile, we also compared PDL1ES to a clinical model (CM; built using clinical characteristics only), a radiomic model (RM; built using only predefined radiomic features), and a deep learning (DL) model (only comprising deep learning features). In addition, we read the machine manufacturer, X-ray tube current and slice thickness from the CT metainformation and then analyzed the influence of these parameters on the PDL1ES results. To verify the robustness of PDL1ES, we randomly re-divided the validation and test cohorts and retrained twice. Furthermore, we used PDL1ES obtained from the model to perform PFS prediction in the immunotherapy dataset. Meanwhile, we also used clinical features to establish a clinical model via Cox regression and performed comparative analysis with respect to PDL1ES.

Statistical analysis

PDL1ES was built in the training cohort and verified in the testing cohort. During the training process, results of the adjustment were primarily determined by the results of the validation cohort. Delong test was used as a method to calculate ROC difference. Moreover, the predicted value of the binarized signature (using X-tile [28]) in predicting the efficacy of immunotherapy was investigated with Kaplan-Meier curves and Log-rank test, which was further compared to the Kaplan-Meier curves obtained based on IHC of PD-L1 expression [28]. We performed all analysis using R software (version 3.5.2; <http://www.R-project.org>) and Python (version 3.6.5,

<https://www.python.org/>).

Results

Clinical characteristics

In the PD-L1 expression dataset, 939 patients were included. The mean age was 58.8 years (± 10.7), and 576 (61.3%) patients were male. There were 432 (46.0%) never-smokers, and PD-L1 expression values were greater than 50% in 328 (34.9%) patients. Demographic and clinical characteristics of the PDL1 expression dataset are shown in Table 1.

In the immunotherapy dataset, 77 patients (81.9%) were male, with a mean age of 60.7 years (± 11.1). Seventy-three (77.7%) were smokers, and the remaining 21 (22.3%) were nonsmokers. Fifty were diagnosed with adenocarcinoma, and the remaining included 33 patients with squamous cell carcinoma, 7 patients with neuroendocrine tumors, 2 patients with lymphoepithelioma-like carcinoma and 2 patients with undifferentiated NSCLC. Fifty-one (54.3%) received pembrolizumab monotherapy as the first-line treatment, and 43 (45.7%) received pembrolizumab combined with chemotherapy as the initial treatment. Clinical characteristics of the immunotherapy dataset are shown in Supplementary Table 1.

Validation of PD-L1 expression signature in predicting PD-L1 expression

We extracted 1316 radiomic features that were calculated from images processed using different filters.

Detailed information about the filters and features is shown in Supplementary Table 2. All radiomic features we extracted abided by the feature definitions that expand on the Imaging Biomarker Standardization Initiative (IBSI) [29].

Figure 2 (A-C) shows the ROC curves of PDL1ES, where the AUCs of the training, validation and test datasets were 0.78 (95% CI: 0.75~0.80), 0.71 (95% CI: 0.60~0.81), and 0.76 (95% CI: 0.67~0.85), respectively. All results were significantly better than the DL, radiomic and clinical models ($P < 0.0001$; ROC test was performed on the PD-L1 expression dataset). We concluded that PDL1ES demonstrated good ability to distinguish between high and low expression of PD-L1 ($\geq 50\%$ and $< 50\%$). In addition, the decision curves of the training and test cohorts further validated our conclusion (Figure 2D-F). We also found that PDL1ES performed well in different layer thicknesses (greater than or equal to 5 mm and less than 5 mm) and X-ray tube current (the median is divided into two queues; cut-off point: 290) (Supplementary Figure 2). In addition, to evaluate the robustness of PDL1ES, we randomly divided the validation and the test cohorts and re-established

the signature. In the experimental verification results of the first random division, the AUC of the training, validation and test cohorts were 0.79 (95% CI: 0.77~0.82), 0.73 (95% CI: 0.61~0.85) and 0.76 (95% CI: 0.67~0.87). In the experimental verification results of the second random division, the AUC of the training, validation and test cohorts were 0.77 (95% CI: 0.74~0.80), 0.74 (95% CI: 0.60~0.86) and 0.75 (95% CI: 0.66~0.84). The results of two additional experiments revealed that PDL1ES was more robust.

Validation of PD-L1 expression signature in predicting the efficacy of immunotherapy

To explore the value of predicting the efficacy of immunotherapy using PD-L1 expression scores, we performed a Kaplan-Meier curve analysis based on PDL1ES. The PDL1ES was binarized with a cutoff of 0.66 obtained with X-tile. The group with PDL1ES greater than 0.66 were termed the high PDL1ES (low-risk) group, while the remaining were termed the low PDL1ES (high-risk) group.

The univariable Cox analysis identified the binarized PDL1ES as a significant prognostic variable (C-index: 0.66, 95%CI: 0.48~0.83; Hazard Ratio (HR): 2.57, 95% CI: 1.22~5.44; $P = 0.010$). In this study, low PDL1ES was associated with improved PFS (Figure 3). The median PFS in the high-risk group was 183 days (95% CI: 122~257) and 363 days (95% CI: 363~) in the low-risk group.

Moreover, stratified analysis was performed based on the treatment regime. For patients who received immunotherapy, the low-risk and high-risk groups showed significant differences in PFS ($P = 0.028$). In patients administered immunotherapy combined with chemotherapy, a median PFS of 271 days was observed in the low-risk group compared to 248 days in the high-risk group ($P = 0.038$).

Exploring the relationship between PD-L1 expression signature and clinical characteristics

We performed a stratified analysis for different clinical characteristics. Among them, we found that except for age ($P_{\text{Age}} = 0.028$), other clinical characteristics were not effective in stratifying patients into high and low-risk ($P_{\text{Family history}} = 0.17$; $P_{\text{Smoking status}} = 0.22$; $P_{\text{Gender}} = 0.95$) groups. The best cut-off value for age was 63. All Kaplan-Meier curves are shown in Supplementary Figure 3.

To further exploit the efficacy of PDL1ES, multivariable Cox regression analysis was performed with a combination of signature and clinical characteristics. Clinical characteristics were not significant when the signature and each characteristic were combined separately ($P_{\text{Family history}} = 0.064$; $P_{\text{Smoking status}} = 0.245$; $P_{\text{Gender}} = 0.374$, $P_{\text{Age}} = 0.088$).

To explore the best predictive level of signature, we used all clinical characteristics to build a clinical model.

Cox regression analysis was performed using a combination of signature and clinical models. We used Cox regression modeling with age, gender, smoking status, and family history. After using X-tile to select the best cutoff point (0.05 in this study), the clinical model classified patients into high-risk and low-risk groups (HR: 2.09, 95% CI: 1.05~4.18; $P = 0.032$) significantly, and the median PFS for high-risk and low-risk is 172 days and 257 days, respectively. Its prognostic stratification effect is not as good as PDL1ES. Using the clinical model in combination with PDL1ES for combinatorial analysis, we found that both variables were very significant ($P_{\text{PDL1ES}} = 0.0396$; $P_{\text{clinical model}} = 0.010$; Cutoff point = 0.05) in the combination of clinical model and PDL1ES. Fortunately, the model also demonstrated strong stratification capabilities (HR: 3.53, 95% CI: 1.86~6.72; $P < 0.001$; Cutoff point = 1.78) and the median PFS has been significantly improved. The median PFS for high and low-risk groups was 122 days and 363 days, respectively. Results of the fusion model were superior to both the clinical and PDL1ES models (Figure 4). Next, we performed the same subgroup analysis of the clinical model and fusion model as PDL1ES. Results indicated that the clinical model was unable to significantly stratify patients in the immunotherapy group and immune combined chemotherapy group ($P_{\text{Immunotherapy}} = 0.068$; $P_{\text{Immunotherapy combined chemotherapy}} = 0.054$), while the fusion model demonstrated better performance than PDL1ES in both groups ($P_{\text{Immunotherapy}} = 0.016$; $P_{\text{Immunotherapy combined chemotherapy}} = 0.003$).

Visualization

A high-expression patient (Patient 1) and a low-expression patient (Patient 2) in the immunotherapy dataset are shown as an example. The class activation map was generated by the gradient of the deep learning to observe important regions that the model considers. At the same time, two representative radiomic features based on univariate analysis were also visualized in the ROI. Details of the patients and the corresponding visualization results are shown in Figure 5.

From this figure, we found that regardless of high and low expression, the primary observations of the deep learning model were located at the border of the lesion. In addition, huge texture differences in the lesion areas between two patients were seen from the visualization of the radiomic features, termed Wavelet-HL_glcM_SumEntropy and Wavelet-LL_glrM_GrayLevelNonUniformityNormalized.

Discussion

Radiomics is an emerging method that can convert medical images into quantitative data to profile tumor phenotypes [30],[31]. In our study, we established PDL1ES via deep learning techniques to identify patients

with high PD-L1 expression with high accuracy ($AUCs \geq 0.71$), which could help to predict the efficacy of immunotherapy in patients with NSCLC.

Currently, IHC is the primary technique utilized to detect PD-L1 expression levels and has many limitations, including sampling bias due to intratumoral heterogeneity and dynamic characteristics of PD-L1 expression levels. As such, some noninvasive PET imaging novel radiotracers, including ^{89}Zr -nivolumab, and ^{18}F -BMS-986192, demonstrated the predictive value of PD-L1 expression. However, they have not been clinically leveraged [32]-[34]. Utilizing the commonly acquired PET/CT images, Jiang et al. assessed PD-L1 expression by radiomic features in NSCLC patients [35], achieving an AUC of 0.80 with only CT radiomic features in predicting PD-L1 (SP142) expression levels over 1%. Though this result is similar to that of our study ($AUC_{\text{training cohort}} = 0.78$; $AUC_{\text{validation cohort}} = 0.71$; $AUC_{\text{test cohort}} = 0.76$), further prognostic validation of the predicted PD-L1 expression was not investigated. There are several benefits in assessing PD-L1 expression levels by radiomic-based signatures. First, its noninvasive manner outperformed IHC which demands tissue samples obtained from surgery or biopsy. Additionally, radiomic-based signatures can dynamically evaluate PD-L1 expression stratification and immune therapeutic efficacy, ultimately informing drug adjustment.

Several studies have evaluated associations between radiomics and other immunotherapy-related genomics or biology. Sun and colleagues evaluated the association between a radiomics-based biomarker of tumor-infiltrating CD8 cells and clinical outcomes of anti-PD-1 and PD-L1 treatment [36]. They found that the gene expression signature of CD8 cells discriminated inflamed tumors from immune-desert tumors ($AUC\ 0.76$). Furthermore, high baseline radiomic scoring was also associated with improved overall survival in patients treated with immunotherapy. Trebeschi and colleagues performed artificial intelligence (AI)-based characterization of primary and metastatic lesions from 203 patients using pretreatment CT imaging data to develop and validate a machine learning biomarker for response to immunotherapy [37]. They found that when combining lesion-wide predictions, immunotherapy response was predicted with an AUC of 0.76 for NSCLC and melanoma ($p < 0.001$). However, various tumor types were included in these studies.

In our study, we first reported on PD-L1 expression prediction using a CT-based deep learning model in NSCLC patients. In addition, we generated visualization of the category activation map and radiomic features. We found that deep learning was more concerned with the marginal zone of the lesion area, and huge textural differences were observed from the visualization maps of texture features, named Wavelet-HL_glcM_SumEntropy and Wavelet-LL_glrM_GrayLevelNonUniformityNormalized. This result confirmed that classification of high and low PD-L1 expression with deep learning features and quantitative radiomic

features were complementary.

Our study has some limitations. First, it was a single-center study, and the predictive value of the obtained imaging biomarker has not been validated in other centers. Second, it was retrospective; therefore, potential bias may exist. In this study, we found that the higher the signature value, the greater the risk, which is contrary to common sense. The reasons for this situation can be summarized into two aspects. First, only a small number of patients was included in the immunotherapy dataset. Secondly, censored data existed in some patients. In addition, since only a minority of patients died in the follow-up period, we did not obtain median OS data. Thus, we are uncertain about the predictive value of this model on long-term efficacy of immunotherapy. In subsequent studies, we will include more immunotherapy patients, and use a combination of retrospective and prospective methods for signature verification and optimization.

In conclusion, the deep learning model provides a noninvasive method for predicting high PD-L1 expression in tumors and for inferring clinical outcomes of immunotherapy. In addition, this deep learning model combined with clinical models improved prediction of the response to immunotherapy.

Declaration of interests

The authors declare no potential conflicts of interest.

Author contributions

WL, ZH and JT conceived and designed the study. DD, BH and WM contributed to data analysis and preparation of the report. KL, LL, YL, HZ, LJ and PZ collected samples and assembled the imaging and clinical data. DD and BH conducted statistical analysis. PT, BH and WM wrote drafted the manuscript. All authors read and approved the final manuscript.

References

- [1]. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018; 68(6), 394-424.
- [2]. Siegel R, DeSantis C and Jemal A. Colorectal cancer statistics, 2014. *CA: a cancer journal for clinicians*. 2014; 64(2), 104-117.
- [3]. Xia L, Liu Y and Wang Y. PD-1/PD-L1 Blockade Therapy in Advanced Non-Small-Cell Lung Cancer: Current

Status and Future Directions. *The oncologist*. 2019; 24(Suppl 1), S31.

- [4]. Chen DS. and Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature*. 2017; 541(7637), 321-330.
- [5]. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nature Reviews Cancer*. 2012; 12(4), 252-264.
- [6]. Atefi M, Avramis E, Lassen A, Wong DJ, Robert L, Foulad D, et al. Effects of MAPK and PI3K pathways on PD-L1 expression in melanoma. *Clinical cancer research*. 2014; 20(13), 3446-3457.
- [7]. Spranger S, Spaapen RM, Zha Y, Williams J, Meng Y, Ha TT, et al. Up-regulation of PD-L1, IDO, and Tregs in the melanoma tumor microenvironment is driven by CD8+ T cells. *Science translational medicine*. 2013; 5(200), 200ra116-200ra116.
- [8]. Bald T, Landsberg J, Lopez-Ramos D, Renn M, Glodde N, Jansen P, et al. Immune cell-poor melanomas benefit from PD-1 blockade after targeted type I IFN activation. *Cancer discovery*. 2014; 4(6), 674-687.
- [9]. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N engl J med*. 2016; 375, 1823-1833.
- [10]. Brody R, Zhang Y, Ballas M, Siddiqui MK, Gupta P, Barker C, et al. PD-L1 expression in advanced NSCLC: Insights into risk stratification and treatment selection from a systematic literature review. *Lung Cancer*. 2017; 112, 200-215.
- [11]. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Updated Analysis of KEYNOTE-024: Pembrolizumab Versus Platinum-Based Chemotherapy for Advanced Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score of 50% or Greater. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2019; 37(7), 537.
- [12]. Mok TS, Wu YL, Kudaba I, Kowalski DM, Cho BC, Turna HZ, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *The Lancet*. 2019; 393(10183), 1819-1830.
- [13]. Scognamiglio G, De Chiara A, Di Bonito M, Tatangelo F, Losito NS, Anniciello A, et al. Variability in immunohistochemical detection of programmed death ligand 1 (PD-L1) in cancer tissue types. *International journal of molecular sciences*. 2016; 17(5), 790.
- [14]. Pinato DJ, Shiner RJ, White SD, Black JR, Trivedi P, Stebbing J, et al. Intra-tumoral heterogeneity in the expression of programmed-death (PD) ligands in isogenic primary and metastatic lung cancer: implications for immunotherapy. *Oncoimmunology*. 2016; 5(9), e1213934.

- [15]. Hofman P. PD-L1 immunohistochemistry for non-small cell lung carcinoma: which strategy should be adopted?. Expert review of molecular diagnostics. 2017; 17(12), 1097-1108..
- [16]. Dong D, Fang MJ, Tang L, Shan XH, Gao JB, Giganti F, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multi-center study. Annals of Oncology. 2020; 31(7), 912-920..
- [17]. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. JAMA oncology. 2016; 2(12), 1636-1642.
- [18]. Song J, Shi J, Dong D, et al, A new approach to predict progression-free survival in stage IV EGFR-mutant NSCLC patients with EGFR-TKI therapy, Clinical Cancer Research, 2018, 24(15): 3583-3592.
- [19]. Khorrami M, Prasanna P, Gupta A, Patil P, Velu PD, Thawani R, et al. Changes in CT Radiomic Features Associated with Lymphocyte Distribution Predict Overall Survival and Response to Immunotherapy in Non–Small Cell Lung Cancer. Cancer Immunology Research, 2020; 8(1), 108-119.
- [20]. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. Theranostics. 2019; 9(5), 1303.
- [21]. Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. Radiotherapy and oncology. 2016; 119(3), 480-486.
- [22]. Kickingereder P, Burth S, Wick A, Götz M, Eidel O, Schlemmer HP, et al. Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. Radiology. 2016; 280(3), 880-889.
- [23]. He B, Dong D, She Y, Zhou C, Fang M, Zhu Y, et al. Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. Journal for immunotherapy of cancer. 2020; 8(2).
- [24]. Zhou H, Dong D, Chen B, et al, Diagnosis of distant metastasis of lung cancer: based on clinical and radiomic features, Translational Oncology, 2018, 11(1): 31-36.
- [25]. Khorrami M, Jain P, Bera K, Alilou M, Thawani R, Patil P, et al. Predicting pathologic response to neoadjuvant chemoradiation in resectable stage III non-small cell lung cancer patients using computed tomography radiomic features. Lung Cancer. 2019; 135, 1-9.
- [26]. Dong D, Zhang F, Zhong LZ, Fang MJ, Huang CL, Yao JJ, et al. Development and validation of a novel MR imaging predictor of response to induction chemotherapy in locoregionally advanced nasopharyngeal cancer: a randomized controlled trial substudy (NCT01245959). BMC medicine. 2019; 17(1), 190.

- [27]. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *The Lancet Oncology*. 2017; 18(3), e143-e152.
- [28]. Camp RL, Dolled-Filhart M and Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clinical cancer research*. 2004; 10(21), 7252-7259.
- [29]. Zwanenburg A, Leger S, Vallières M and Löck S. Image biomarker standardisation initiative; 2016. ArXiv161207003 Cs. 2019.
- [30]. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Annals of Oncology*. 2019; 30(3), 431-438.
- [31]. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Velazquez ER, Parmar C, et al. Defining the biological basis of radiomic phenotypes in lung cancer. *Elife*. 2017; 6, e23421.
- [32]. Donnelly DJ, Smith RA, Morin P, Lipovšek D, Gokemeijer J, Cohen D, et al. Synthesis and biologic evaluation of a novel ¹⁸F-labeled adnectin as a PET radioligand for imaging PD-L1 expression. *Journal of Nuclear Medicine*. 2018; 59(3), 529-535.
- [33]. Cole EL, Kim J, Donnelly DJ, Smith RA, Cohen D, Lafont V, et al. Radiosynthesis and preclinical PET evaluation of ⁸⁹Zr-nivolumab (BMS-936558) in healthy non-human primates. *Bioorganic & medicinal chemistry*. 2017; 25(20), 5407-5414.
- [34]. Niemeijer AN, Leung D, Huisman MC, Bahce I, Hoekstra OS, Van Dongen GAMS, et al. Whole body PD-1 and PD-L1 positron emission tomography in patients with non-small-cell lung cancer. *Nature communications*. 2018; 9(1), 1-5.
- [35]. Jiang M, Sun D, Guo Y, Guo Y, Xiao J, Wang L, et al. Assessing PD-L1 expression level by radiomic features from PET/CT in nonsmall cell lung cancer patients: an initial result. *Academic Radiology*. 2020; 27(2), 171-179.
- [36]. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *The Lancet Oncology*. 2018; 19(9), 1180-1191.
- [37]. Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Călin AM, Delli Pizzi A, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Annals of Oncology*. 2019; 30(6), 998-1004.

Figures

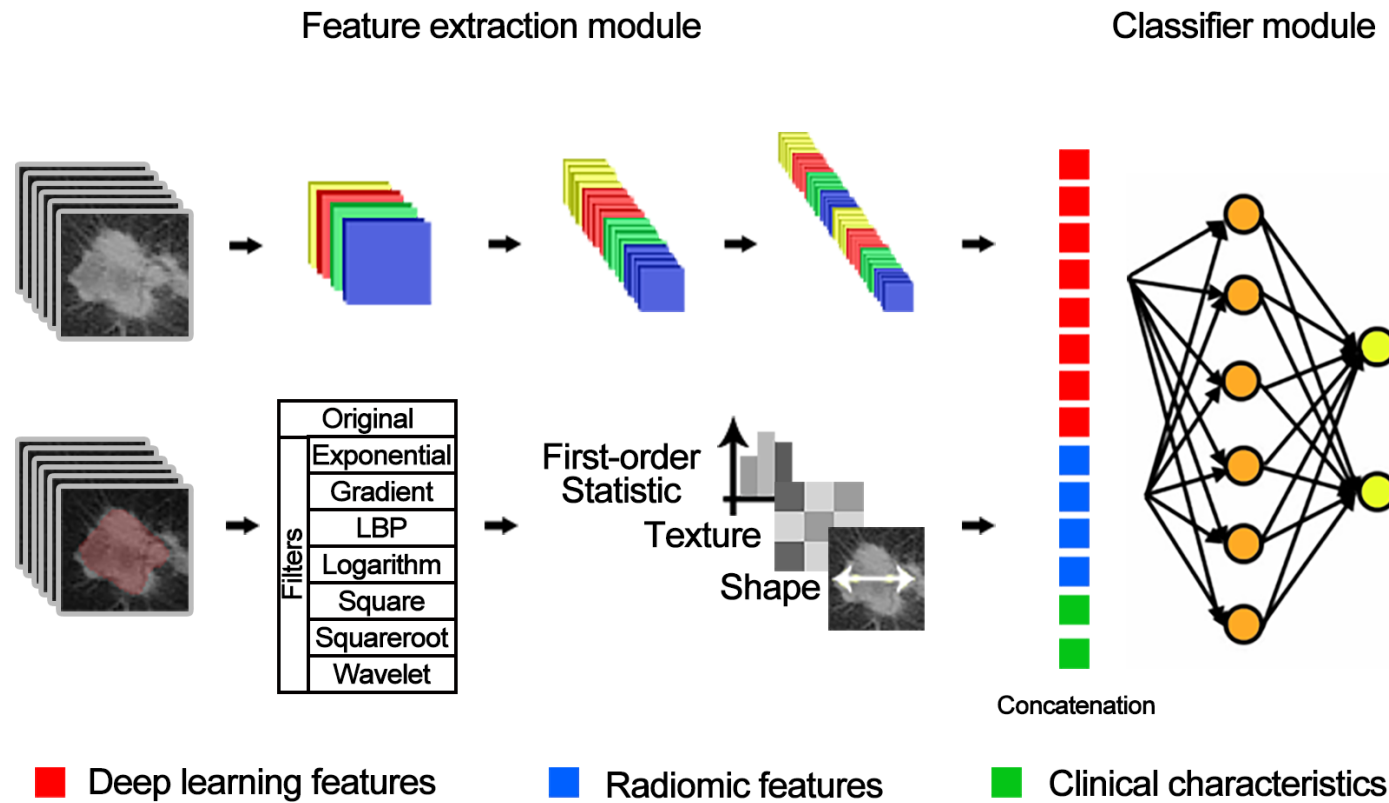


Figure 1. The structure of deep learning network. The entire network includes a feature extraction module (a convolutional network) for extracting deep learning features, and a classification module (a fully-connected network) for classifying PDL1 expressions based on deep learning features combined with predefined radiomic features and clinical characteristics.

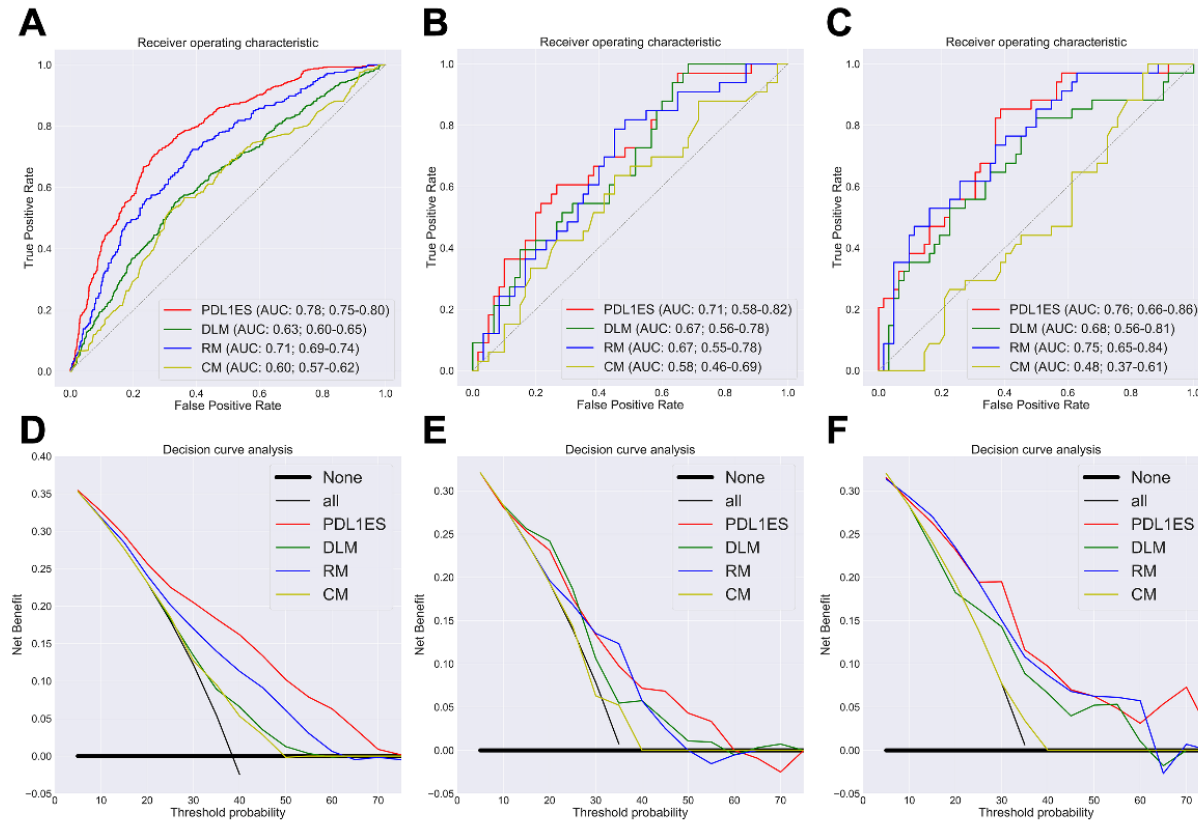


Figure 2. PDL1 expression signature assessment results for high and low expression classification. (A), (B) and (C) were the ROC curves of the training, validation, and test cohorts, respectively. (D), (E) and (F) were the decision curves of the training, validation, and test cohorts, respectively. Among them, each figure contained the results of PDL1 expression signature (PDL1ES), deep learning model (DLM), radiomic model (RM) and clinical model (CM). They all showed that PDL1 expression signature was superior to other models.

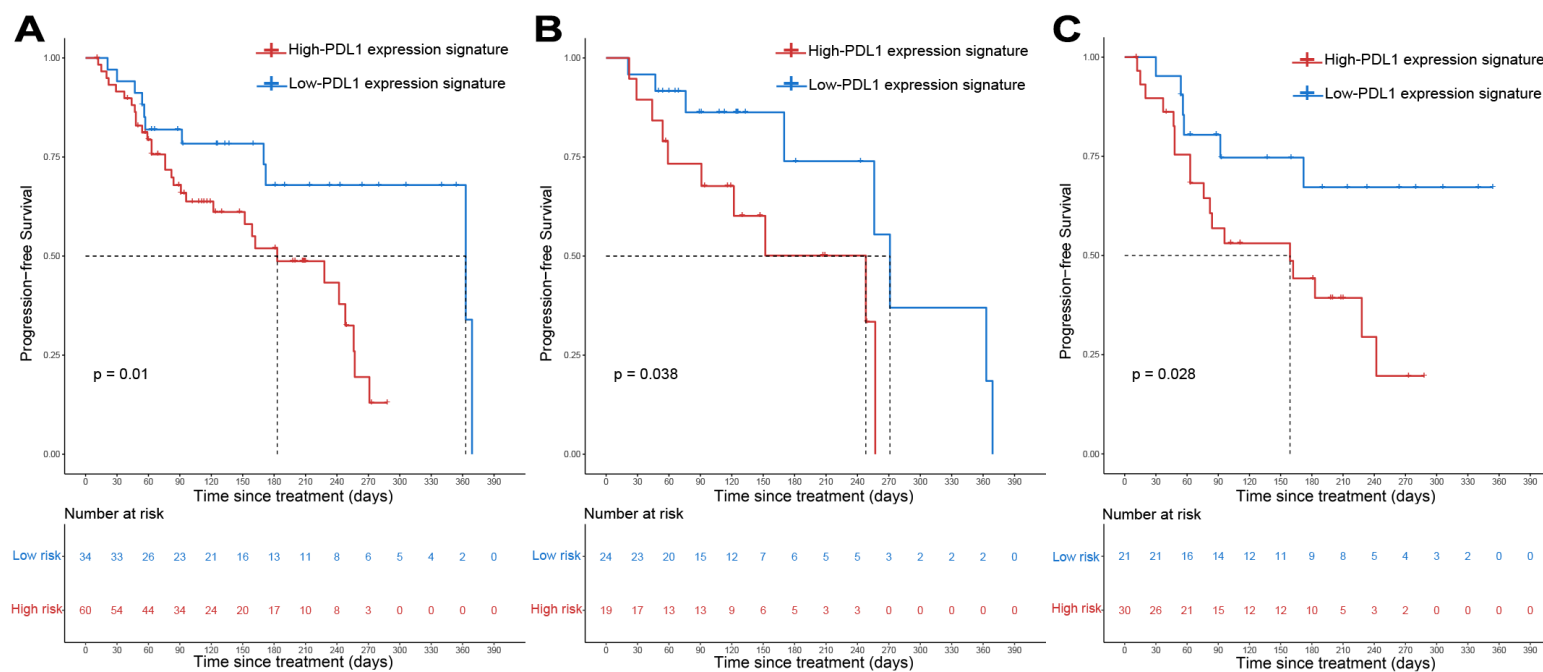


Figure 3. Prognostic performance in different subgroups, comparing high PDL1 expression signature and low PDL1 expression signature. A. Progression-free survival (PFS) of patients relative to PDL1 expression signature (high or low, as defined by the median value) in whole immunotherapy dataset. B. PFS of patients in pembrolizumab combined with chemotherapy subgroup. C. PFS of patients in pembrolizumab-monotherapy subgroup.

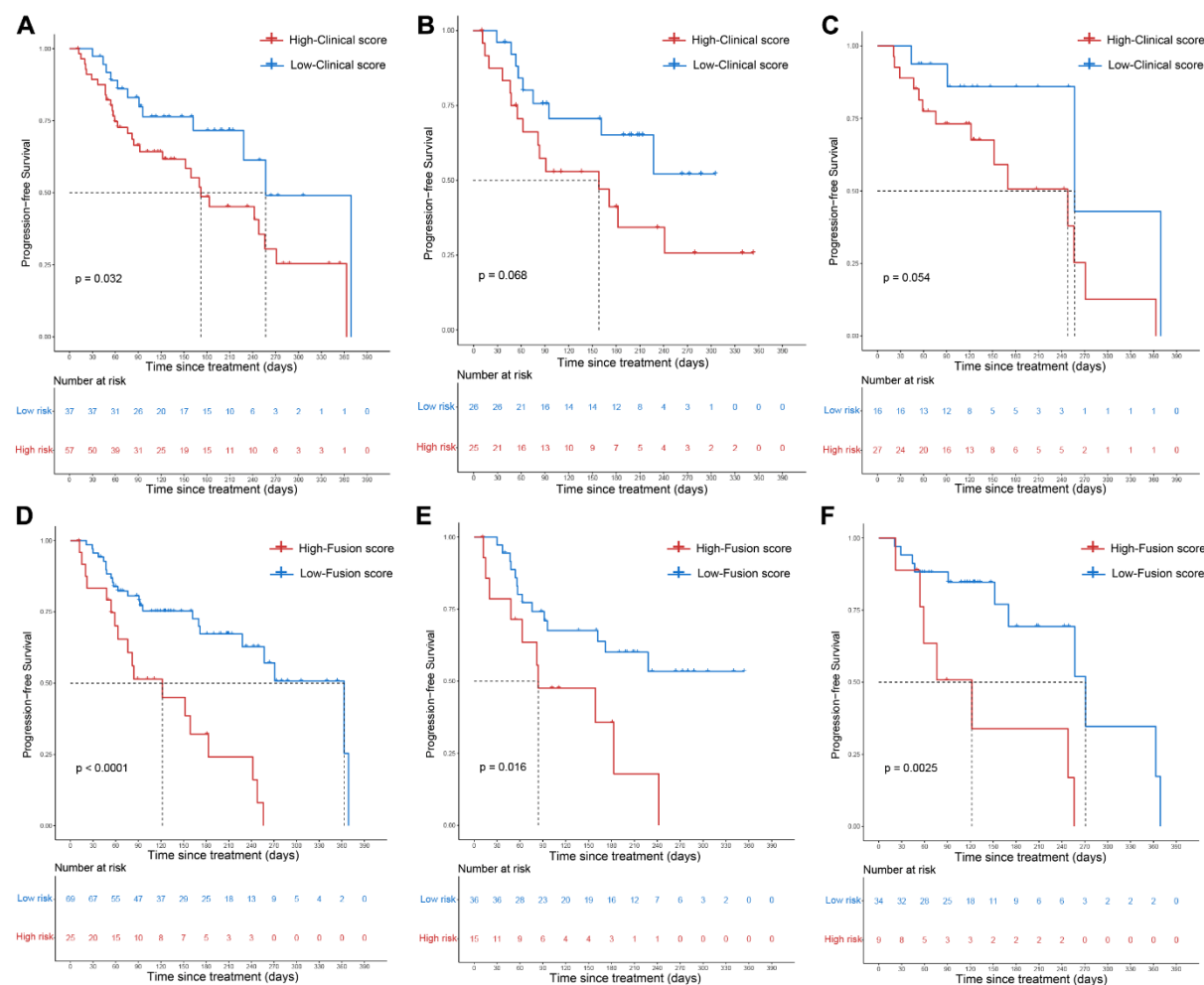


Figure 4. Prognostic performance of clinical model and fusion model. (A), (B) and (C) were the KM curves which indicated progression-free survival (PFS) of patients relative to clinical model in entire immunotherapy cohort, pembrolizumab-monotherapy subgroup and pembrolizumab combined with chemotherapy subgroup, respectively. (D), (E) and (F) were the KM curves which indicated PFS of patients relative to combination of PDL1 expression signature (PDL1ES) and clinical model in entire immunotherapy cohort, pembrolizumab-monotherapy subgroup and pembrolizumab combined with chemotherapy subgroup, respectively.

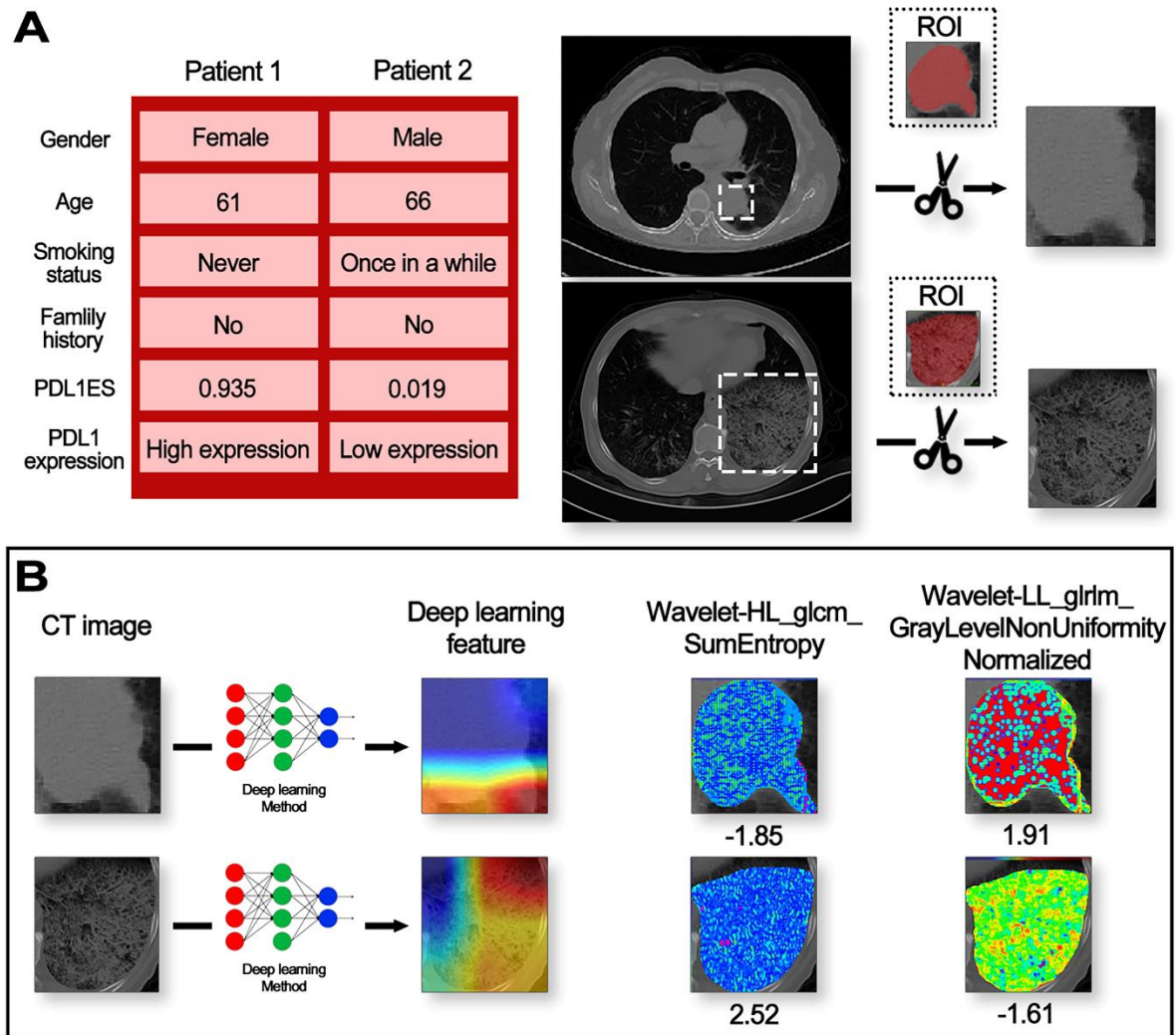


Figure 5. The information of patients and model feature visualization. (A) contained a random sample of a high-expression patient and a low-expression patient and their corresponding CT images. (B) indicated the visual images of the imaging features (including deep learning features and typical radiomic features).

Tables

Table 1 Clinical characteristics of the PDL1 expression dataset

Characteristics	Total (n=939)	Training cohort (n=750)	Validation cohort (n=93)	<i>p</i> Value*	Test cohort (n=96)	<i>p</i> Value**
Age, year, mean±SD	58.8 ± 10.7	58.9 ± 10.7	58.0 ± 10.3	0.462	59.0 ± 11.3	0.892
Gender, n (%)				0.838		0.408
Male	576 (61.3)	455 (60.7)	58 (62.4)		63 (65.6)	
Female	363 (38.7)	295 (39.3)	35 (37.6)		33 (34.4)	
Smoking status, n (%)				0.938		0.409
Never	432 (46.0)	348 (46.4)	43 (46.2)		41 (42.7)	
Occasionally	374 (39.8)	300 (40.0)	36 (38.7)		38 (39.6)	
Constantly	95 (10.1)	74 (9.7)	11 (11.8)		10 (10.4)	
Unknown	38 (4.0)	28 (3.7)	3 (3.2)		7 (7.2)	
Family history of cancer, n (%)				0.686		0.437
No	764 (81.4)	610 (81.3)	79 (84.9)		75 (78.1)	
Yes	139 (14.8)	113 (15.1)	11 (11.8)		15 (15.6)	
Unknown	36 (3.8)	27 (3.6)	3 (3.2)		6 (6.3)	
PD-L1 expression, n (%)				0.988		0.995
≥ 50%	328 (34.9)	261 (34.8)	33 (35.5)		34 (35.4)	
< 50%	611 (65.1)	489 (65.2)	60 (64.5)		62 (64.6)	

Categorical data are shown as numbers (%) and continuous data as mean ± SD;

* the p Value is the test result of the training cohort and the validation cohort;

** the p Value is the test result of the training cohort and the test cohort.