World Journal of Surgical Oncology

REVIEW

Open Access

Predictive value of machine learning for PD-L1 expression in NSCLC: a systematic review and meta-analysis

Ting Zheng[1*], Xingxing Li[1], Li Zhou[1] and Jianjiang Jin[1]

Abstract

Background As machine learning (ML) continuously develops in cancer diagnosis and treatment, some researchers have attempted to predict the expression of programmed death ligand-1 (PD-L1) in non-small cell lung cancer (NSCLC) by ML. However, there is a lack of systematic evidence on the effectiveness of ML.

Methods We conducted a thorough search across Embase, PubMed, the Cochrane Library, and Web of Science from inception to December 14th, 2023.A systematic review and meta-analysis was conducted to assess the value of ML for predicting PD-L1 expression in NSCLC.

Results Totally 30 studies with 12,898 NSCLC patients were included. The thresholds of PD-L1 expression level were < 1%, 1–49%, and ≥ 50%. In the validation set, in the binary classification for PD-L1 ≥ 1%, the pooled C-index was 0.646 (95%CI: 0.587–0.705), 0.799 (95%CI: 0.782–0.817), 0.806 (95%CI: 0.753–0.858), and 0.800 (95%CI: 0.717–0.883), respectively, for the clinical feature-, radiomics-, radiomics + clinical feature-, and pathomics-based ML models; in the binary classification for PD-L1 ≥ 50%, the pooled C-index was 0.649 (95%CI: 0.553–0.744), 0.771 (95%CI: 0.728–0.814), and 0.826 (95%CI: 0.783–0.869), respectively, for the clinical feature-, radiomics-, and radiomics + clinical feature-based ML models.

Conclusions At present, radiomics- or pathomics-based ML methods are applied for the prediction of PD-L1 expression in NSCLC, which both achieve satisfactory accuracy. In particular, the radiomics-based ML method seems to have wider clinical applicability as a non-invasive diagnostic tool. Both radiomics and pathomics serve as processing methods for medical images. In the future, we expect to develop medical image-based DL methods for intelligently predicting PD-L1 expression.

Keywords Lung cancer, PD-L1, Machine learning, Meta-analysis, Radiomics

*Correspondence:
Ting Zheng
zz_tt666@163.com
[1]Department of Medical Oncology, The First People's Hospital of Linping District, Hangzhou 311100, Zhejiang Province, China

Introduction

With an estimated 2.2 million new cases and 1.8 million fatalities in 2020, lung cancer is the second most frequent disease worldwide and the major cause of cancer-related deaths [1]. 80–85% of primary lung cancer cases are non-small cell lung cancer (NSCLC) [2], and it is treated primarily with surgery, chemotherapy, radiotherapy, and targeted therapy.

The advent of immunotherapy in recent years has fundamentally changed the paradigm for treating advanced NSCLC and altered the prognosis of early-stage NSCLC [3]. Studies have shown that immunotherapy prolongs both progression-free survival (PFS).

and overall survival (OS) in advanced NSCLC patients without driver gene mutations [4–6]. The expression level of PD-L1 is seemingly a predictor for the response of NSCLC patients to immunotherapy. According to National Comprehensive Cancer Network (NCCN) guidelines, single-agent pembrolizumab, atezolizumab, or cemiplimab-rwlc is recommended (Category 1; preferred) as first-line therapy, with a median OS of 20 months for metastatic NSCLC patients lacking targetable driver mutations and with PD-L1 ≥ 50% (KEYNOTE-042, IMpower110, and EMPOWER-Lung1) [7, 8]. Single-agent pembrolizumab is recommended (Category 2B; useful in certain circumstances) as first-line therapy for patients with metastatic NSCLC regardless of medical history, PD-L1 = 1-49%, and negative for actionable driver mutations [7]; it is also recommended (Category 1; preferred) as subsequent therapy for patients with metastatic NSCLC and PD-L ≥ 1% without undergoing immunotherapy [6]. Atezolizumab is proposed as an adjuvant therapy for patients with completely resected (R0) stage IIA-IIIB NSCLC, PD-L1 ≥ 1%, and negative for driver mutations (IMpower010) [9]. Immunotherapy plus chemotherapy is recommended (Category 1; preferred) as first-line therapy for patients with metastatic NSCLC and negative for actionable driver mutations, regardless of the PD-L1 expression. Many studies have shown that greater clinical benefits can be yielded from immunotherapy in case of a higher PD-L1 expression [10–12].

As recommended by the NCCN guidelines, immunohistochemistry (IHC) testing is applied to the prediction of PD-L1 expression. However, tumor tissue should be harvested for IHC testing, which is invasive and time-consuming, failing to dynamically reflect the PD-L1 expression [13]. Moreover, it is impossible to predict the PD-L1 expression by too few cells in biopsy samples. In clinical practice, therefore, developing a new efficient, low-cost, accurate, and rapid method for predicting PD-L1 is essential.

As artificial intelligence (AI) continuously develops, its efficiency in cancer diagnosis and treatment has been increasingly verified [14–16]. Chen et al. verified in a systematic review and meta-analysis that machine learning (ML) is accurate in preoperative prediction of genetic mutations in lung cancer [17]. Didier et al. also proved that ML models are promising in predicting OS in lung cancer patients [18]. In recent years, some researchers have attempted to apply ML to predict PD-L1 expression in lung cancer. In particular, radiomics-based ML is a non-invasive imaging method that plays a non-negligible role in cancer diagnosis and clinical management [19] as it can convert imaging data into mineable data and extract a variety of high-throughput quantitative information from medical images by automatic or semi-automatic analysis. In the context of diverse ML methods and models, the application value of ML is inconclusive. Therefore, the predictive value of ML for PD-L1 expression in NSCLC was assessed in this paper. Therefore, this systematic review and meta-analysis was conducted on the predictive value of ML for the PD-L1 expression in NSCLC patients to provide an evidence-based rationale for the development of artificial intelligence in this field.

Methods

Study registration

This systematic review and meta-analysis was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-analyses(PRISMA) [20], and registered with PROSPERO (CRD42024504947)(https://www.crd.york.ac.uk/PROSPERO/view/CRD42024504947).

Search strategy

We systematically searched English-language studies in Embase, PubMed, Web of Science, and the Cochrane Library from inception to December 14th, 2023. Medical subject headings plus free words were used in the search without restriction on dates and regions, including "lung cancer", "machine learning", and "Programmed death-ligand 1". The search procedures and strategies are shown in Table S1.

Eligibility criteria

Inclusion criteria

(1) Patients diagnosed histologically with primary NSCLC.
(2) Cohort studies, cross-sectional studies, case-control studies, and clinical trials.
(3) ML models were established for predicting the PD-L1 expression.
(4) English-language studies.

Exclusion criteria

(1) Meta-analyses, expert opinions, guidelines, and reviews, etc.

(2) Complete ML models were not established with only differential factor analysis.
(3) Absence of the following outcome measures affecting the predictive accuracy of ML models: receiver operating characteristic (ROC), confusion matrix, c-statistic, accuracy, calibration curve, diagnostic fourfold table, sensitivity, C-index, specificity, recovery, and F1 score.
(4) A too-small sample size (< 20 cases).

Study selection and data extraction

First, we imported the literature retrieved into EndNote. Titles and abstracts were examined to filter the studies after duplicates were eliminated. Then the full texts were examined to ultimately determine the eligible studies.

A spreadsheet was created before data extraction to record the following data: first author, publication year, country, patient source, tumor stage, threshold of PD-L1 expression, radiomics source, Image area of Interest (ROI) region segmentation software, Total number of all outcome event cases, Total number of cases, number of cases in training set and validation set, method for generation of validation set, model types and modeling variables.

Two investigators (TZ and XXL) were independently responsible for screening all studies and extracting data. In case of disagreement, they could consult a third investigator for resolution.

Risk of bias (RoB) in studies

The RoB in the original studies included was assessed using the PROBAST [21], which reflected the overall RoB and overall applicability. PROBAST contained questions in four domains: participants (two specific questions), predictors (three specific questions), outcomes (six specific questions), and statistical analysis (nine specific questions), each with three responses (Yes/Probably yes, No/Probably no, and No information). The risk of a domain was deemed high if the answer to at least one question in this domain was "No" or "Probably no", and low if the answer to all questions was "Yes" or "Probably yes". The overall RoB could be considered low only when the risk of all domains was low, and considered higher when the risk of at least one domain was high.

Two investigators were independently responsible for assessing RoB by the PROBAST and cross-checking. If there was a disagreement, they could consult a third investigator for resolution.

Pooled methods

C-index that assesses the overall accuracy of ML models underwent a meta-analysis. If the C-index had no 95%CI and standard error in some of the original studies, the standard error was estimated with reference to the study by Debray TP et al. [22]. Due to the variations in the variables and parameters of ML models, the meta-analysis was performed by a random-effects model.

In addition, sensitivity and specificity underwent a meta-analysis by a bivariate mixed-effects model using a diagnostic fourfold table. However, the diagnostic fourfold table was not reported in most original studies, which was calculated using specificity, sensitivity, precision combined with the number of cases. Stata15.1 was used for the meta-analysis.

Results

Study selection

A total of 1578 studies were retrieved from the four databases, of which 593 were duplicate publications and 951 were excluded after reading titles and abstracts. After the full texts of the remaining 34 studies were reviewed, four studies that lacked outcome measured were eliminated. Finally, 30 studies were included (Fig. 1).

Study characteristics

The 30 studies included were published in 2019–2023, involving 12,898 NSCLC patients. They were all case-control studies including six multicenter studies [23–28]. Regarding the tumor stage, the patients were mostly in stage I-IV [23–36], one study on early stage I-II [37], and seven on mid-late stage III-IV [23, 37–42]. The threshold of PD-L1 expression was ≥ 1% in 12 studies [23, 27–32, 36, 43–46], ≥ 50% in four studies [38, 39, 42, 47], and both ≥ 1% and ≥ 50% in 14 studies [24–26, 33–35, 37, 40, 41, 48–52]. Radiomics- and pathomics-derived variables were used for modeling, with 23 studies based on radiomics models [23, 27–36, 38–42, 44–48, 51, 52], and seven studies based on pathomics models [24–26, 37, 43, 49, 50]. Among the studies based on radiomics-based models, seven studies were derived from positron emission tomography/computed tomography (PET/CT) [27, 29, 30, 35, 45, 51, 52], 15 studies from computed tomography (CT) [28, 31–34, 36, 38–42, 44, 46–48], and one study from magnetic resonance imaging (MRI) [23]. The validation set was generated by external validation in one study [23], and by random sampling in 17 studies [27–34, 36, 38–40, 44, 47, 48, 51, 52]. The detailed basic information is displayed in Tables 1 and 2.

RoB in studies

All studies included were case-control studies, but deep learning (DL) was adopted in 27 models. Due to less impact of case-control studies on DL, DL models in case-control studies were assessed as low RoB. Certain RoB may be brought by case-control studies to the assessment of traditional ML models, so the traditional ML models were assessed as high RoB. The prediction of PD-L1 expression mainly relied on IHC, and these

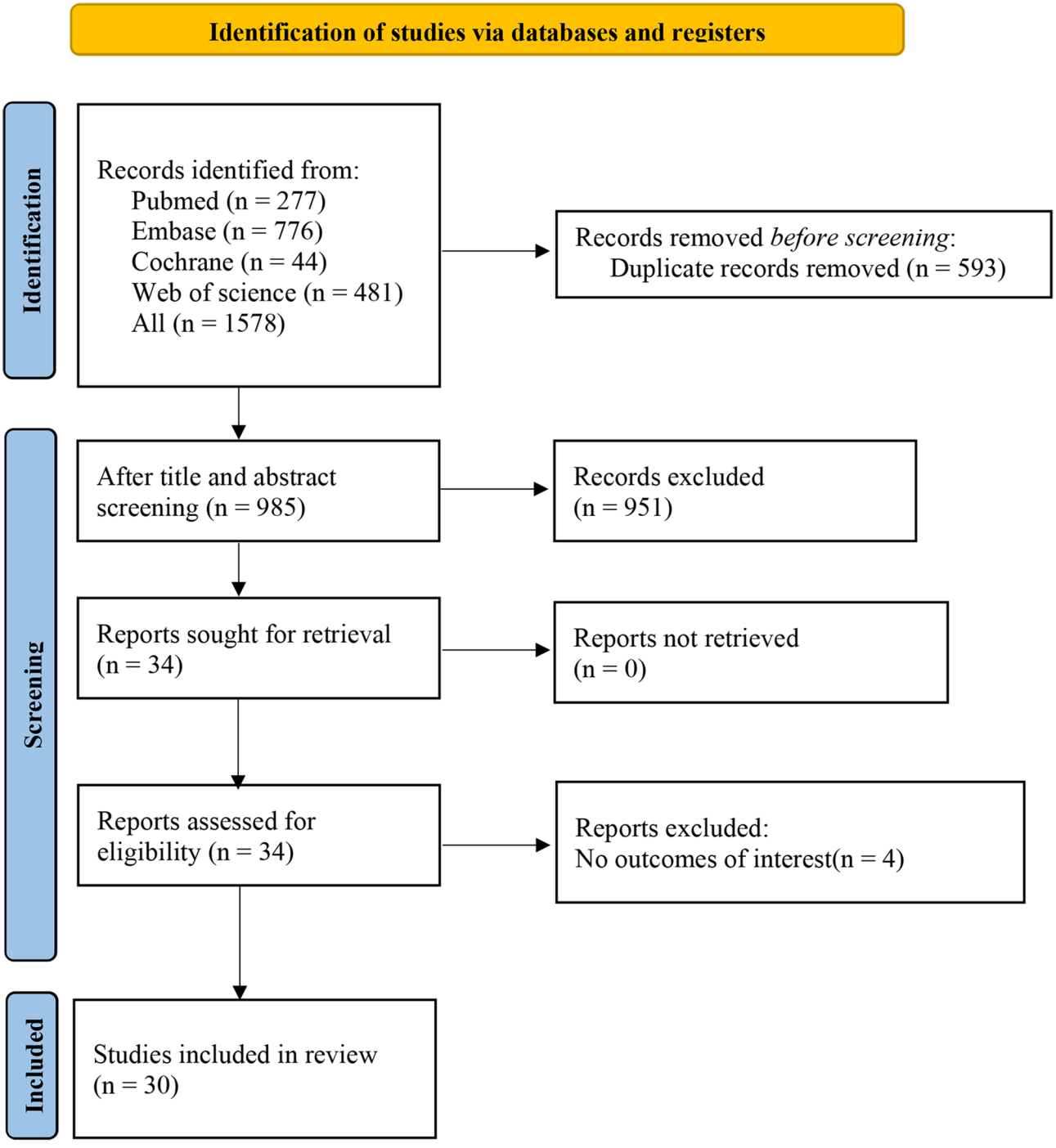

Fig. 1 Literature screening procedure

modeling variables had no impact on the results, so the RoB was low. In the statistical analysis, the events per variable (EPV) ≥ 20 in 40 models caused high RoB. EPV of another 78 models could not be calculated, but radiomics or pathomics features were used, so the RoB could not be calculated and thus was assessed as unclear; whether overfitting assessment was performed was not reported in 37 models, so the RoB was high. The RoB in studies is detailed in Fig. 2.

Meta-analysis

Results of binary classification.

Table 1 Basic information of included studies employing radiomics

No.	First author	Year of publication	Country	Patient source	tumor stage	PD-L1 expression threshold	Radiomics source	Image area of Interest (ROI) region segmentation software	Total number of all outcome event cases	Total number of cases	Number of cases in training set	Generation of validation set	Number of cases in the validation set	Model type	Modeling variables
1	Xiaoqian Zhao	2023	china	Single center	I-IV	≥1%	PET/CT	LIFEx	162	334	233	Random sampling	101	LR	radiomic, clinical, radiomic + clinical
2	Y.B. Wang	2023	china	Single center	I-IV	≥1%	PET/CT	ITK-SNAP	159	394	275	Random sampling	119	LR	radiomic, radiomic + clinical
3	Anna-Katharina Meißner	2023	Germany	Multiple centers	IV	≥1%	MRI	ITK-SNAP	25	53	36	External validation	17	RF	radiomic, clinical, radiomic + clinical
4	P.M. Liu	2023	China	Single center	I-IV	≥1%	CT		68	125	74	Random sampling	51	DL	radiomic, clinical
5	Kohei Hashimoto	2023	Japan	Single center	I/II	≥1%	CT	Eclipse software	33	117	79	Random sampling	38	RF	radiomic, clinical, radiomic + clinical
6	Yu Fu	2023	china	Single center		≥1%	CT		37	54				ANN, SVM	radiomic
7	Ruiyun Zhang	2022	Germany	Single center		≥1%	PET/CT	LIFEx	32	58				LR	radiomic
8	Chengdi Wang	2022	china	Single center	I-IV	≥1%	CT		590	1691	698	Random sampling	Test set 175 Validation set 818	DL	radiomic, radiomic + clinical
9	Chengdi Wang	2022	china	Single center	I-IV	≥1%, ≥50%	CT		1–49% 50 ≥ 50% 363	1135	908	Random sampling	227	DL	radiomic
10	Takehiro Shiinoki	2022	Japan	Single center	I-IV	≥1%, ≥50%	CT	a 3D slicer	≥1% 98 ≥50% 36	161	112	Random sampling	49	LightGBM	radiomic, clinical
11	Jun Shao	2022	china	Single center		≥1%, ≥50%	CT		300	843	675	Random sampling	Validation set 84 Test set 84	DL	radiomic
12	CHAE HONG LIM	2022	Korea	Single center	I-IV	≥1%, ≥50%	PET/CT	LIFEx 4.0	≥1% 154 ≥50% 61	312				RF, ANN,NB, LR, AdaBoost, SGD,SVM	radiomic, clinical, radiomic + clinical
13	Qiang Wen	2021	china	Single center	III–IV	≥50%	CT	a 3D Slicer	77	120	90	Random sampling	30	LR	radiomic, clinical, radiomic + clinical
14	Chengdi Wang	2021	china	Single center	I-IV	≥1%	CT	ITK-SNAP	484	1262	882	Random sampling	Validation set 125 Test set 225	DL	radiomic
15	Panwen Tian	2021	china	Single center	IV	≥50%	CT		328	939	750	Random sampling	Validation set 93 Test set 96	DL	radiomic, clinical
16	Wei Mu	2021	USA	Multiple centers	I-IV	≥1%	PET/CT	ITK-SNAP	168	485	284	Random sampling	Validation set 116 Test set 85	DL	radiomic, clinical
17	Jihui Li	2021	china	Single center	I-IV	≥1%, ≥50%	PET/CT	LIFEx	≥1% 154 ≥50% 69	255	170	Random sampling	85	LR	radiomic, clinical, radiomic + clinical
18	Zekun Jiang	2021	china	Multiple centers	I-IV	≥1%	CT	ITK-SNAP	89	125	91	Random sampling	34	RF, DT, LR, AdaBoost, Gaussian process, SVM	radiomic, clinical, radiomic + clinical
19	Stefano Bracci	2021	Italy	Single center	IIIA–IV	≥1%, ≥50%	CT	LIFEx	≥1% 49 ≥50% 24	72	48	Random sampling	24	LR	radiomic
20	Ying Zhu	2020	china	Single center	III–IV	≥1%, ≥50%	CT	ITK-SNAP	≥1% 46 ≥50% 38	127				DL, LR	radiomic
21	Jiyoung Yoon	2020	South Korea	Single center	III–IV	≥50%	CT	AVIEW Research	53	153				LR	radiomic, clinical, radiomic + clinical
22	Zongqiong Sun	2020	china	Single center	I-IV	≥50%	CT	ITK-snap	185	390	260	Random sampling	130	LR	radiomic, clinical, radiomic + clinical
23	Mengmeng Jiang	2019	china	Single center	I-IV	≥1%, ≥50%	PET/CT	ITK-snap	≥1% 265 ≥50% 90	399	266	Random sampling	133	LR, RF	radiomic

LR: logistic regression, RF: random forest, DL: deep learning, MLP: multi-layer perceptron, SVM: Support Vector Machine, ANN: Neural Network, NB: Naïve Bayes, AdaBoost: Adaptive Boosting, DT: decision tree, SGD: stochastic gradient descent

PD-L1 ≥ 1%

Pooled results

In the validation set, in the binary classification for PD-L1 ≥ 1%, the pooled C-index was 0.646 (95%CI: 0.587–0.705), 0.799 (95%CI: 0.782–0.817), 0.806 (95%CI: 0.753–0.858), and 0.800 (95%CI: 0.717–0.883) for the clinical feature-, radiomics-, radiomics + clinical feature-, and pathomics-based ML models, with pooled sensitivity and specificity of 0.62(95%CI:0.45–0.77) and 0.62 (95%CI:0.55–0.69), 0.75(95%CI:0.70–0.79)and 0.78(95%CI:0.73–0.83), 0.75 (95%CI:0.69–0.80) and 0.76 (95%CI:0.67–0.84), and 0.76–0.95 and 0.76–0.97, respectively (Table 3).

Subgroup analysis

The results of subgroup analyses showed that in the validation set, the pooled C-index was 0.811 (95%CI: 0.778–0.845), 0.700 (95%CI: 0.555–0.846), 0.800 (95%CI: 0.766–0.834), and 0.760 (95%CI: 0.544–0.975) for the CT-, PET-, PET-CT-, and MRI-based radiomics models, with pooled sensitivity and specificity of 0.78 (95%CI:0.71–0.83) and 0.80 (95%CI:0.73–0.86), 0.63–0.81 and 0.52–0.71, 0.71 (95%CI:0.66–0.76) and 0.80 (95%CI:0.75–0.84), and 0.71 and 0.80, respectively (Table 3).

Among the radiomics + clinical feature-based models, in the validation set, the CT-, positron emission tomography-(PET-), PET-CT-, and MRI-based models had a pooled C-index of 0.819 (95%CI: 0.742–0.896), 0.806 (95%CI: 0.801–0.810), 0.765 (95%CI: 0.703–0.827), and 0.840 (95%CI: 0.659–1.021), respectively (Table 3).

Reporting biases

The funnel plot of ML for predicting PD-L1 ≥ 1% showed that a significant publication bias might arise in the validation set (Egger's test: $P < 0.05$) (Fig. 3).

PD-L1 ≥ 50%

Pooled results

In the validation set, in the binary classification for PD-L1 ≥ 50%, the pooled C-index was 0.649 (95%CI: 0.553–0.744), 0.771 (95%CI: 0.728–0.814), and 0.826 (95%CI: 0.783–0.869) for the clinical feature-, radiomics-, and radiomics + clinical feature-based ML models, with pooled sensitivity and specificity of 0.73 (95%CI:0.59–0.83) and 0.59 (95%CI:0.46–0.70), 0.75 (95%CI:0.70–0.78) and 0.72 (95%CI:0.66–0.78), and 0.80–0.89 and 0.50–0.72, respectively. The pathomics-based ML model had pooled sensitivity and specificity of 0.75 and 0.99, respectively (Table 4).

Subgroup analysis

The results of subgroup analyses revealed that in the validation set, the pooled C-index was 0.769 (95%CI: 0.718–0.820), 0.715 (95%CI: 0.629-0.800), and 0.829 (95%CI: 0.715–0.944) for the CT-, PET-, and PET-CT-based radiomics models, with pooled sensitivity and specificity of 0.75 (95%CI:0.70–0.80) and 0.71 (95%CI:0.65–0.76), 0.62–0.72 and 0.62–0.71, and 0.72–0.77 and 0.62–0.92, respectively (Table 4).

Among the radiomics + clinical feature-based models, in the validation set, the pooled C-index was 0.828 (95%CI: 0.776–0.880) and 0.814 (95%CI:0.715–0.913), respectively, for the CT- and PET-CT-based models (Table 4).

Reporting biases

The funnel plot of ML for predicting PD-L1 ≥ 50% showed that a significant publication bias might arise in the validation set (Egger's test: $P < 0.05$). In contrast, the clinical feature-based ML model had no significant publication bias (Egger's test: $P > 0.05$) (Fig. 4).

Results of multiclass classification

Multiclass classification was described in five studies [25, 26, 33, 37, 50], which mainly identified negative PD-L1, PD-L1 = 1–49%, and PD-L1 ≥ 50%. The meta-analysis results revealed that the predictive accuracy for negative PD-L1, PD-L1 = 1–49%, and PD-L1 ≥ 50% was 74.4% (95%CI: 46.4–94.7) (Fig. 5), 85.3% (95%CI: 71.9–95.2) (Fig. 6), and 85.7% (95%CI: 81.6–89.4) (Fig. 7), respectively.

Discussion

This systematic review showed that radiomics and pathomics analyses were the main methods for predicting PD-L1 expression. In radiomics, medical images were mainly from CT, PET-CT, and MRI, especially the first two, whose good predictive power had been verified. Nonetheless, it is important to acknowledge the significance of the clinical features. In the validation set, in the binary classification for PD-L1 ≥ 1%, the pooled C-index was 0.799 (95%CI: 0.782–0.817) and 0.806 (95%CI: 0.753–0.858) for the radiomics- and radiomics + clinical feature-based ML models, with pooled sensitivity and specificity of 0.75 (95%CI:0.70–0.79) and 0.78 (95%CI:0.73–0.83), and 0.75 (95%CI:0.69–0.80) and 0.76 (95%CI:0.67–0.84), respectively.

Seol et al. used radiomics features derived from PET-CT to predict PD-L1 expression [53], and found that the area under the summary ROC curve was 0.74, similar to the finding in this paper. They did not, however, take clinical features' influence on radiomics models' prediction power into account quantitatively, and they only included a small number of primary studies. In contrast, the clinical features were taken into account in this paper, and it was found that the combination of PET-CT and clinical

Table 2 Basic information of included studies employing pathomics

No.	First author	Year of publication	Country	Patient source	tumor stage	PD-L1 expression threshold	Number of overall outcome events	Total number of cases	Number of cases in training set	Generation of validation set	Number of cases in validation set	Model type
24	Jianghua Wu	2021	china	Multiple centers		≥ 1%, ≥ 50%				Random sampling		DL
25	Liesbeth M Hondelink	2022	Netherlands	Single center	IV	≥ 1%, ≥ 50%	1–49% 56, ≥ 50% 46	199(22C3)	60	Random sampling	139	DL
26	Sangjoon Choi	2022	Korea	Multiple centers		≥ 1%, ≥ 50%		1281(22C3)	802	External validation	479	DL
27	Guoping Cheng	2022	china	Single center		≥ 1%, ≥ 50%	22C3: 1–50% 255, ≥ 50% 191, SP263: 1–50% 10, ≥ 50% 11	1288(22C3: 1204, SP263: 84)	627(22C3)	Random sampling	Validation set (22C3) (*N* = 577)Validation set (SP263)(*N* = 84)	DL
28	Xiangyun Wang	2021	china	Multiple centers		≥ 1%, ≥ 50%		300(22C3)	190	Random sampling	110	DL
29	Jingxin Liu	2021	china	Single center		≥ 1%, ≥ 50%	1–49% 28 ≥ 50% 51	96	45	Random sampling	51	DL
30	Lingdao Sha	2019	USA	Single center	I- IV	≥ 1%	69	130	48	Random sampling	82	DL

DL: deep learning

features produced a better C-index. (PD-L1 ≥ 1%: 0.765; PD-L1 ≥ 50%: 0.814).

Previously, the desirable accuracy of radiomics for predicting immunotherapy response and outcome in patients with NSCLC had also been verified. Chen et al. found in a meta-analysis [54] that the sensitivity and specificity of radiomics for predicting immunotherapy response and outcome of NSCLC are 76% and 84%, respectively. This study applied ML to the prediction of PD-L1 expression before the treatment of NSCLC, which, as early screening prior to immunotherapy, also obtained relatively favorable results.

In clinical practice, modeling variables are a key factor to improving the predictive value of the ML model. In this paper, it was found that the ML modeling variables for the prediction of PD-L1 expression in NSCLC were mainly clinical features, radiomics, radiomics + clinical features, and pathomics. Radiomics-based ML, as a non-invasive prediction means, has attracted extensive attention in the field of lung cancer prediction [17, 18] In addition, radiomics-based ML exhibits high accuracy in the prediction of PD-L1 expression in NSCLC. In contrast, pathomics-based ML is an invasive prediction means, possibly with a desirable accuracy which remains to be further enhanced. Since it is difficult to predict the PD-L1 expression by common clinical features, the predictive value of ML based only on clinical features is limited for PD-L1 expression.

For radiomics-based ML for the prediction of PD-L1 in NSCLC, the medical images are derived primarily from CT, and also PET-CT and MRI. CT is the major prediction means for early diagnosis and clinical staging of lung cancer, whereas PET-CT and MRI are not essential ones. In clinical practice, PET-CT and MRI are mainly used to assess extrapulmonary metastases. However, the widespread clinical application of PET-CT is restricted due to high cost. Therefore, efforts should be made to develop efficient CT-based prediction methods in the future. In the included studies, the primary methods include radiomics approaches based on CT, MRI, or PET-CT, which have demonstrated promising predictive performance.

In recent years, DL methods have also been gradually paid extensive attention by researchers. For traditional ML methods, demarcation and encoding of ROI in an image are required. Encoding is carried out using intelligent software, but human-computer interaction is needed for demarcation, so the researcher's priori knowledge will bring about a certain bias to a large extent. In contrast, DL based on image processing can be trained relying on raw unprocessed images, which avoids the impact of priori knowledge to a certain extent. In this study, it was found that DL methods possibly had superior discriminative ability to traditional ML methods. In

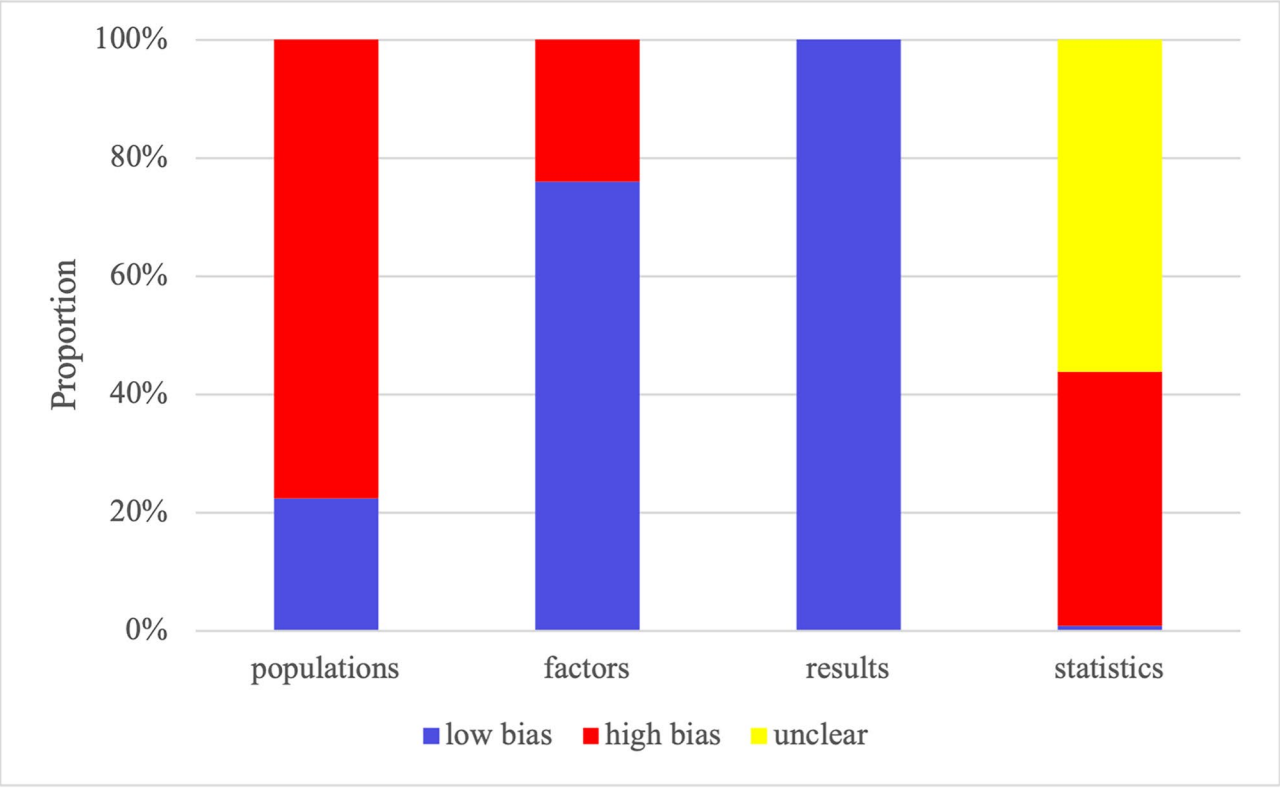

Fig. 2 Results of RoB assessment in included studies

the future, therefore, we can actively try to develop DL models with better prediction performance to raise the PD-L1 prediction accuracy.

The models included in this study were predominantly based on binary classification, whereas multiclass classification is often more applicable in clinical practice. Immunotherapy alone is recommended as the first-line therapy for NSCLC patients lacking targetable driver gene mutations and with PD-L1 levels ≥ 50% [7, 8]. In addition, patients with PD-L1 = 1–49% [10, 55] and those without any significant PD-L1 expression [8, 56] have an improved response rate to immunotherapy with checkpoint inhibitors. To sum up, greater clinical benefits can be yielded from immunotherapy in case of a higher PD-L1 expression. Therefore, it is believed that the multiclass classification model is more clinically applicable. However, the binary classification model is dominant currently. The small number of multiclass classification studies included had preferable values for the prediction of negative PD-L1, PD-L1 = 1–49%, and PD-L1 ≥ 50%.

In this study, some models had high RoB due to small sample sizes or a lack of external validation. The following several reasons can explain the high RoB. Firstly, the RoB tool considers non-prospective studies or databases to be at high RoB, and the included studies were mostly single-center case-control studies, so they were assessed as high RoB since case-control studies might introduce some bias in the assessment of modeling variables and interpretation of results. Moreover, case-control studies from public non-databases also caused high risk in predictors. Secondly, the RoB of statistical analyses was mainly due to EPV < 20 in the training set or a lack of a validation set of more than 100 cases, which is a very strict rule for the current model. For small-scale studies, however, it is difficult to satisfy the condition that the number of positive events in the training set is over 10 times larger than that ultimately incorporated into the model or an independent validation set of more than 100 cases is available. In addition, EPV cannot be calculated accurately for image-based models because image-based models, especially deep learning, in the original study had no concept of modeling variables, and the researchers using the image-based traditional ML were often unwilling to or did not report the detailed image parameters included finally. Therefore, the RoB tool is a very strict tool assessing the included studies, and it seemingly should be updated to a greater extent in future studies.

While pooled performance metrics are promising, the heterogeneity among included studies is substantial. Although we have tried our best to conduct subgroup analysis according to different radiomics and different machine learning algorithms, there is still great heterogeneity. This may be because in the process of radiomics, image segmentation may be dependent on different

Table 3 Subgroup analysis of c-index, sensitivity and specificity of machine learning in the prediction of PD-L1 ≥ 1%

subgroup	c-index					sensitivity and specificity		
	n	events	samplesize	c-index(95%CI)	I^2	*n*	sensitivity(95%CI)	specificity(95%CI)
Clinical features	8	273	542	0.646(0.587–0.705)	39.2	7	0.62(0.45–0.77)	0.62(0.55–0.69)
Radiomics								
CT								
DL	13	590	1606	0.827(0.798–0.850)	53.9	13	0.76(0.67–0.83)	0.84(0.77–0.89)
other ML	7	276	482	0.783(0.731–0.835)	70	6	0.80(0.72–0.85)	0.69(0.58–0.79)
overall	20	866	2088	0.811(0.778–0.845)	86.5	19	0.78(0.71–0.83)	0.80(0.73–0.86)
PET								
other ML	3	187	337	0.700(0.555–0.846)	94.4	3	0.63–0.81	0.52–0.71
overall	3	187	337	0.700(0.555–0.846)	94.4	3	0.63–0.81	0.52–0.71
PET/CT								
DL	3	107	230	0.828(0.776–0.880)	0	3	0.69–0.83	0.60–0.89
other ML	4	241	438	0.789(0.746–0.833)	64	4	0.69(0.63–0.75)	0.78(0.71–0.83)
overall	7	348	668	0.800(0.766–0.834)	52.3	7	0.71(0.66–0.76)	0.80(0.75–0.84)
MRI								
other ML	1	7	17	0.760(0.544–0.975)	NA	1	0.71	0.80
overall	1	7	17	0.760(0.544–0.975)	NA	1	0.71	0.80
Overall	31	1034	2436	0.799(0.782–0.817)	92.8	30	0.75(0.70–0.79)	0.78(0.73–0.83)
Radiomics + Clinical features								
CT	4	381	1065	0.819(0.742–0.896)	85.8	4	0.79(0.71–0.85)	0.80(0.64–0.90)
PET	1	48	119	0.806(0.801–0.810)	NA	1	0.79	0.69
PET/CT	2	105	186	0.765(0.703–0.827)	0	2	0.60–0.69	0.68–0.77
MRI	1	7	17	0.840(0.659–1.021)	NA	1	1	0.70
Overall	8	541	1387	0.806(0.753–0.858)	92.4	8	0.75(0.69–0.80)	0.76(0.67–0.84)
Pathological features	1	41	82	0.800(0.717–0.883)	NA	3	0.76–0.95	0.76–0.97

DL: deep learning; ML: machine learning

imaging devices and image parameters, and is also limited by the clinical experience of the segmenter during the segmentation process. In addition, there may be certain differences in the predictive performance between different machine learning models. Therefore, more standardized guidelines should be developed for radiomics to enhance its transparency and promote its clinical application.

Limitations

This paper provides for the first time an evidence-based rationale for the value of ML for the prediction of PD-L1 expression. However, this study still had some limitations. Firstly, several radiomics features were covered in the included studies, but such studies were insufficient, which may restrict the interpretation of the results to a certain extent. Secondly, only DL and traditional ML were differentiated due to an insufficient number of studies, and the predictive accuracy was not explored across ML methods in detail. Thirdly, very few multiclass classification studies were included, so in the future, more multiclass classification studies are required to validate the findings. Fourthly, many of the included studies are single-center or regional, considering the global relevance of NSCLC, which might introduce bias due to localized patient characteristics or treatment protocols, restricting the interpretation of results. Fifthly, model complexity, hyperparameter tuning, and feature engineering all significantly influence model performance. Since applied studies were included in this paper, and they involved no model complexity, hyperparameter tuning, or feature engineering, we were unable to summarize these parameters in our study, which was also a limitation. Sixth, QUADAS-2 is mainly used in randomized diagnostic experiments. Among randomized diagnostic experiments, case-control studies are considered to have a greater risk of bias. The studies included in our analysis are mainly single-center retrospective case-control studies, which may cause a high risk of bias. This is also a challenge in machine learning research, and it is a limitation of our study. Seventh, there seems to be a regional concentration in the included studies. Healthcare practices and procedures may vary across different regions, which may limit the generalizability and applicability of the research results. However, due to the limited number of the included studies, we were unable to deeply discuss its impact on the results among different task types. At the same time, multi-center cross-border studies are desired in the future to develop artificial intelligence detection tools that cover a wider range of information.

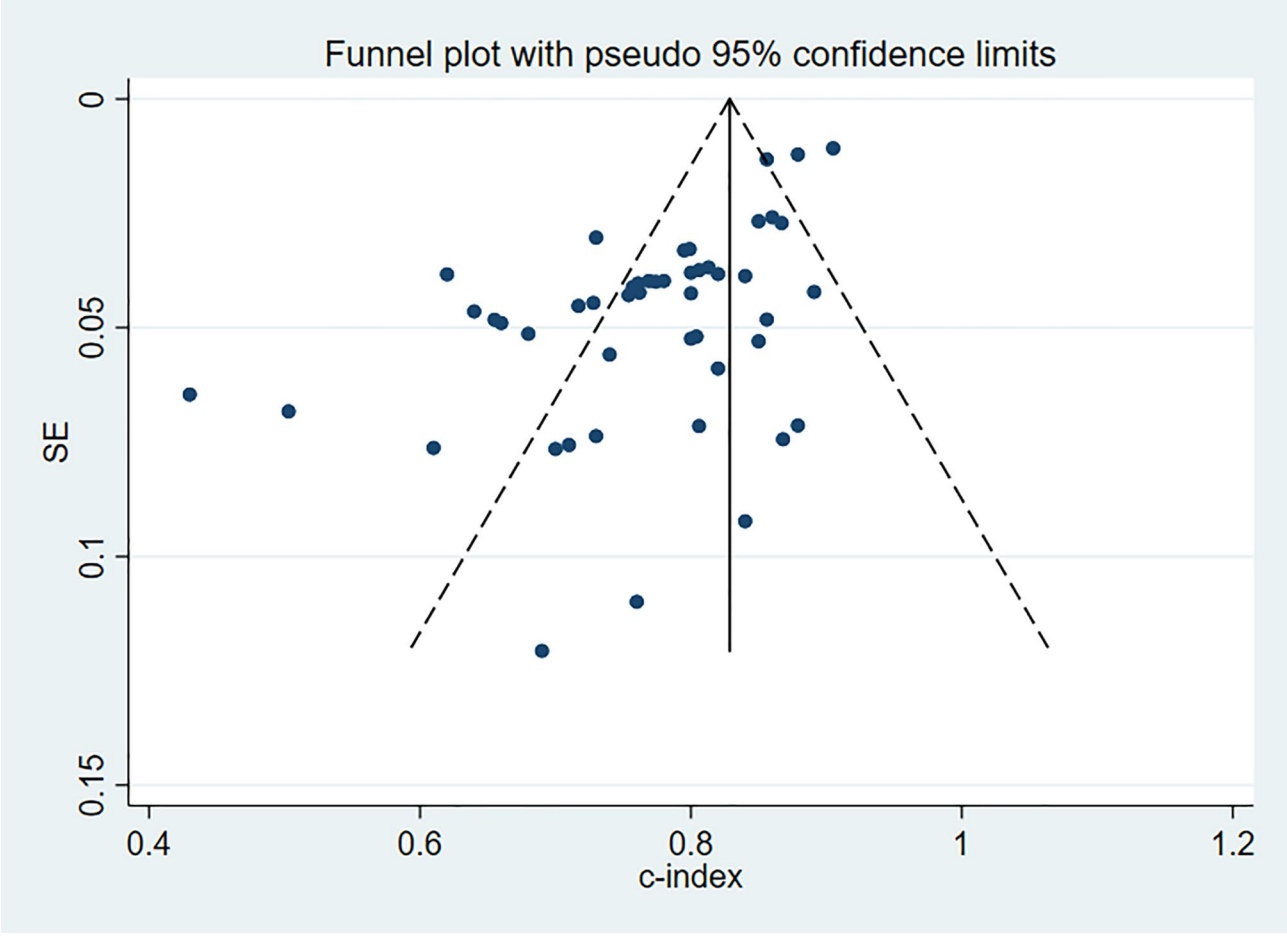

Fig. 3 Funnel plot of ML for predicting PD-L1 ≥ 1%

Table 4 Subgroup analysis of c-index, sensitivity and specificity of machine learning in the prediction of PD-L1 ≥ 50%

subgroup	c-index					sensitivity and specificity		
	n	events	samplesize	c-index(95%CI)	I^2	*n*	sensitivity(95%CI)	specificity(95%CI)
Clinical features	6	182	483	0.649(0.553–0.744)	74.3	5	0.73(0.59–0.83)	0.59(0.46–0.70)
Radiomics								
CT								
DL	5	105	316	0.737(0.692–0.782)	0	5	0.75(0.66–0.82)	0.65(0.58–0.71)
other ML	5	142	402	0.803(0.724–0.881)	72.1	5	0.77(0.69–0.83)	0.78(0.72–0.82)
overall	10	247	718	0.769(0.718–0.820)	64.8	10	0.75(0.70–0.80)	0.71(0.65–0.76)
PET								
other ML	2	53	218	0.715(0.629-0.800)	17.7	2	0.62–0.72	0.62–0.71
overall	2	53	218	0.715(0.629-0.800)	17.7	2	0.62–0.72	0.62–0.71
PET/CT								
other ML	2	53	218	0.829(0.715–0.944)	68.5	2	0.72–0.77	0.62–0.92
overall	2	53	218	0.829(0.715–0.944)	68.5	2	0.72–0.77	0.62–0.92
Overall	14	247	718	0.771(0.728–0.814)	64.9	14	0.75(0.70–0.78)	0.72(0.66–0.78)
Radiomics + Clinical features								
CT	2	81	160	0.828(0.776–0.880)	13.6	2	0.83–0.89	0.50–0.72
PET/CT	1	23	85	0.814(0.715–0.913)	NA	1	0.80	0.70
Overall	3	104	245	0.826(0.783–0.869)	0	3	0.80–0.89	0.50–0.72
Pathological features						1	0.75	0.99

DL: deep learning; ML: machine learning

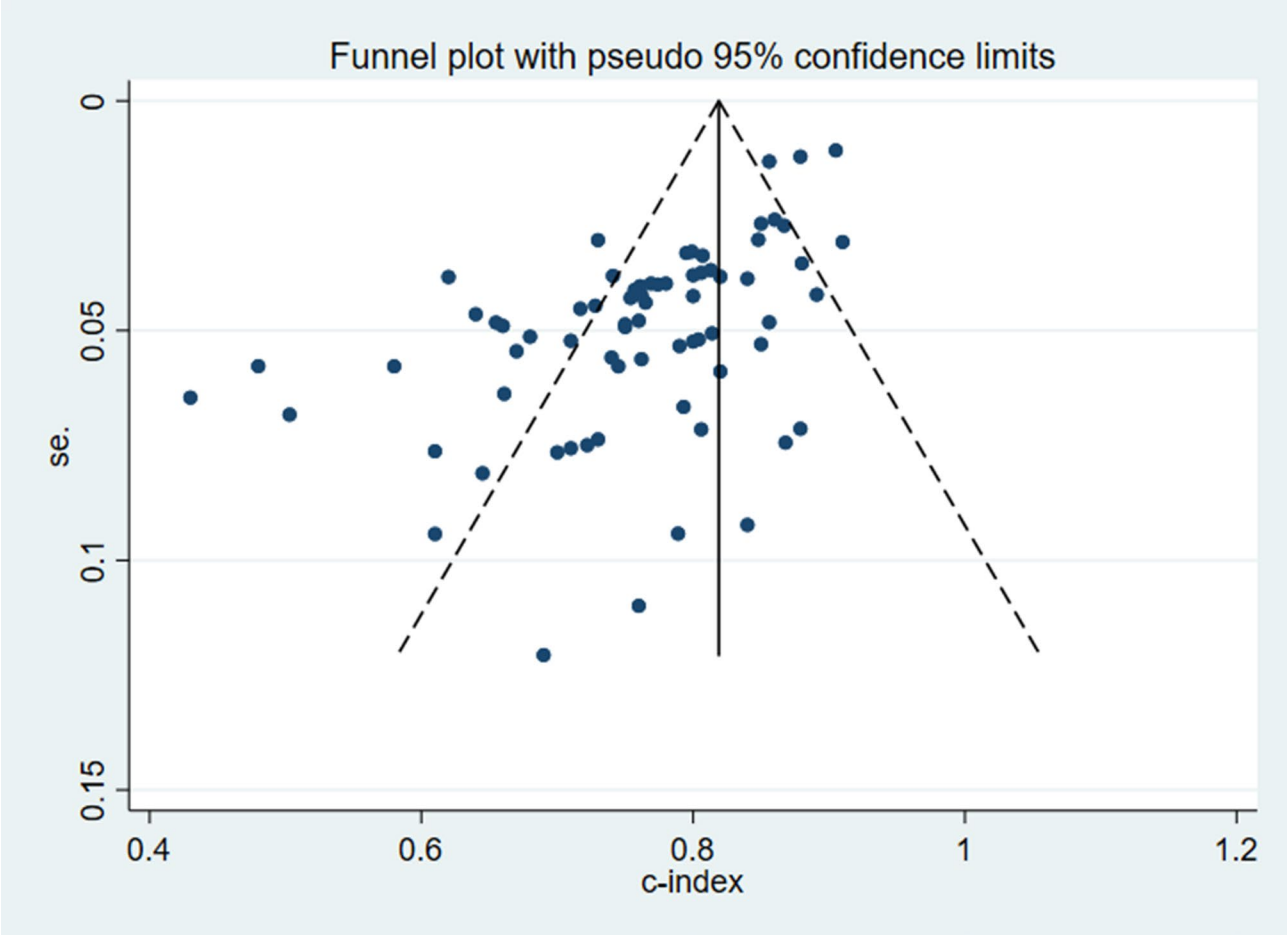

Fig. 4 Funnel plot of ML for predicting PD-L1 ≥ 50%

Future perspective

This study demonstrated that machine learning possesses a good predictive performance for the PD-L1 expression. The included studies were predominantly based on binary classification, and the different expression levels of PD-L1 should be considered in clinical practice. In future studies, therefore, efficient regression models should be established based on multiclass classification or accurate expression for the prediction of PD-L1 expression. In addition, the performance of ML is mainly verified by random sampling in the existing studies, which is a serious challenge for radiomics studies. Therefore, the accuracy of models should be validated in a multicenter large-sample study in future studies. Meanwhile, few studies are available on deep learning, and the current studies mainly focus on ML. ML can achieve intelligent image processing. Hence, it may also be a focus of work in the future.

Conclusions

This study demonstrates that ML methods, especially radiomics-based ML, achieve more desirable accuracy for the prediction of PD-L1 expression in NSCLC, and deep learning seemingly exhibits better prediction performance. The included studies in this paper were predominantly based on binary classification, but negative PD-L1, PD-L1 = 1–49%, and PD-L1 ≥ 50% should be taken into account in clinical practice. Therefore, the value of deep learning for the prediction of PD-L1 expression in NSCLC should be more deeply explored by multiclass classification models in the future.

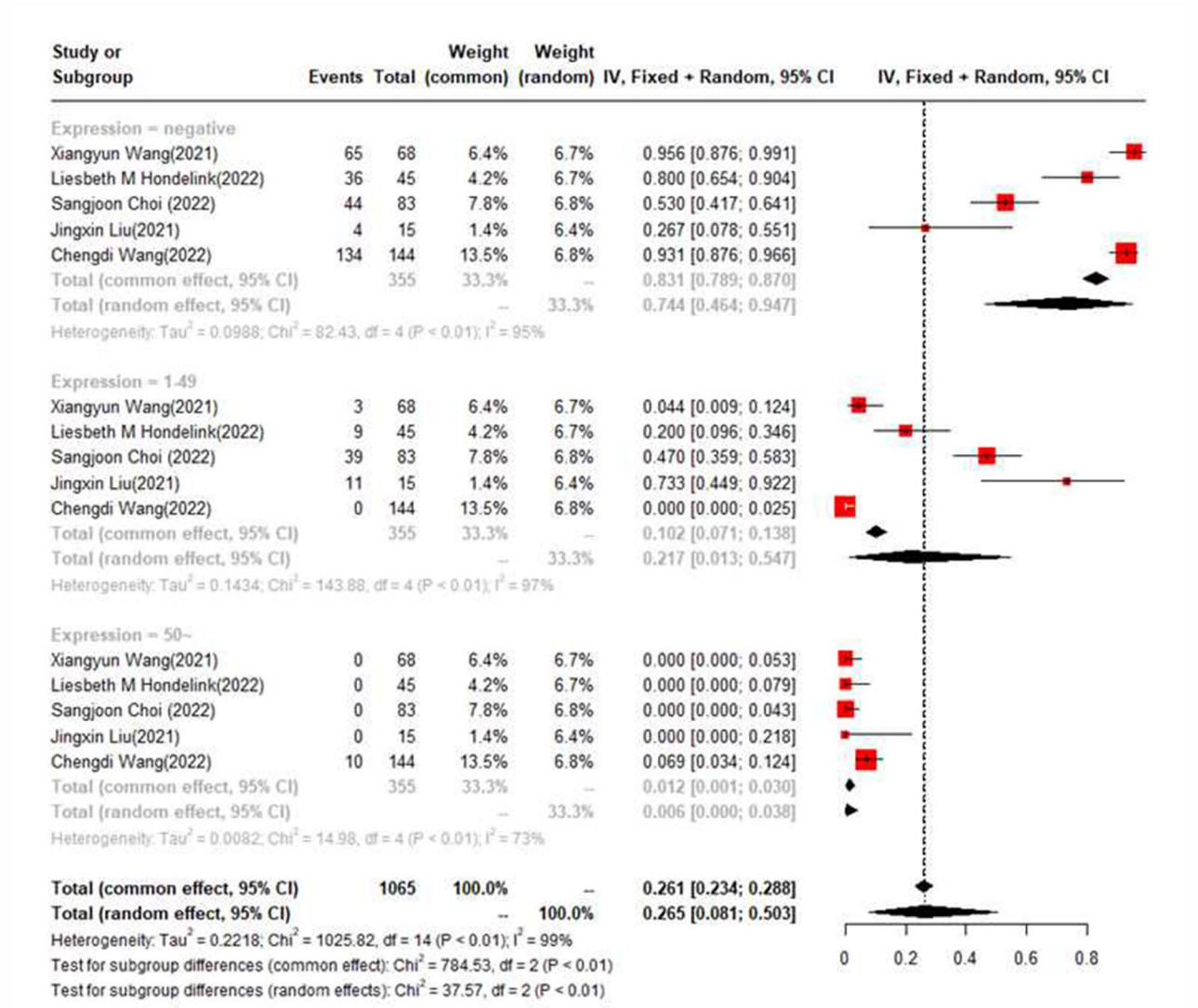

Fig. 5 Forest plot of accuracy of machine learning in the prediction of negative PD-L1 expression with multiclass classification

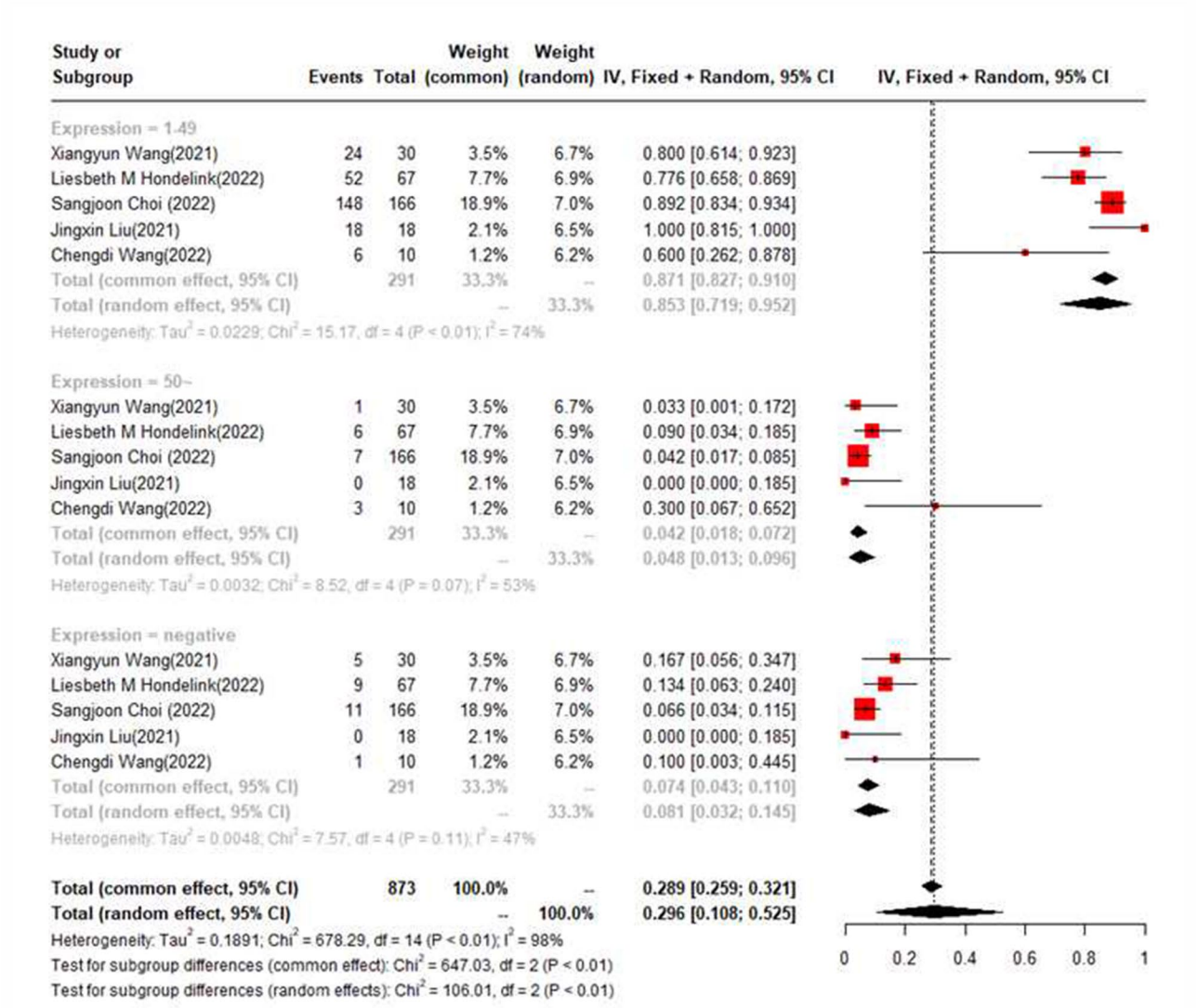

Fig. 6 Forest plot of accuracy of machine learning in the prediction of PD-L1 = 1–49% expression with multiclass classification

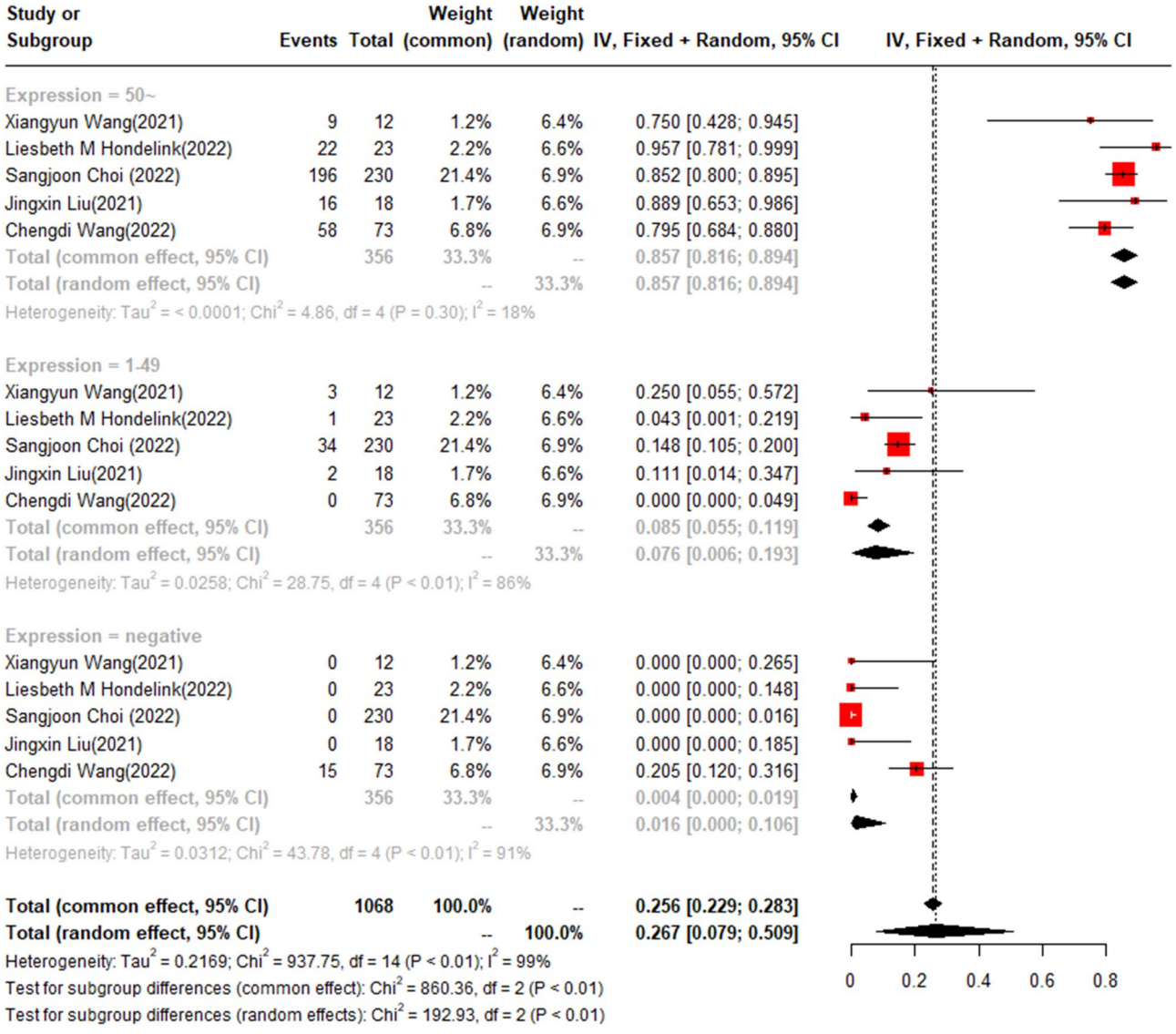

Fig. 7 Forest plot of accuracy of machine learning in the prediction of PD-L1 ≥ 50% expression with multiclass classification

Abbreviations

NSCLC	Non-small cell lung cancer
PD-L1	Programmed cell death ligand-1
ML	Machine learning
PFS	Progression-free survival
OS	Overall survival
ROC	Receiver operating characteristic
NCCN	National Comprehensive Cancer Network
IHC	Immunohistochemistry
AI	Artificial intelligence
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analyses
ROI	Image area of Interest
RoB	Risk of bias
EPV	Events per variable
PET/CT	Positron emission tomography/computed tomography
CT	Computed tomography
MRI	Magnetic resonance imaging

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12957-025-03847-6.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
Not applicable.

Author contributions
All authors contributed to the study conception and design. Writing - original draft preparation: Ting Zheng; Writing - review and editing: Ting Zheng, Xingxing Li; Conceptualization: Ting Zheng; Methodology: Ting Zheng, Li Zhou; Formal analysis and investigation: Ting Zheng, Xingxing Li, Li Zhou, Jianjiang Jin; Supervision: Ting Zheng, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by Medical Health Technology Project of Hangzhou (grant number B20230153), Clinical Medicine Special Fund Project of Zhejiang Medical Association (grant number 2024ZYC-A294), Lingping District Science and Technology Planning Project(grant number LPWJ 2024-02-28) and Traditional Chinese Medicine Science and Technology Project of Zhejiang Province (grant number 2023ZR128).

Data availability
All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval
Not applicable.

Consent to participate
Not applicable.

Consent to publication
Not applicable.

Received: 24 February 2025 / Accepted: 11 May 2025
Published online: 22 May 2025

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Siegel R, Desantis C, Jemal A. Colorectal cancer statistics, 2014. CA Cancer J Clin. 2014;64(2):104–17.
3. Deslypere G, Gullentops D, Wauters E, Vansteenkiste J. Immunotherapy in non-metastatic non-small cell lung cancer: can the benefits of stage IV therapy be translated into earlier stages? Ther Adv Med Oncol. 2018;10:1758835918772810.
4. Rizvi NA, Mazières J, Planchard D, Stinchcombe TE, Dy GK, Antonia SJ, et al. Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial. Lancet Oncol. 2015;16(3):257–65.
5. 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(19):1823–33.
6. Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet. 2016;387(10027):1540–50.
7. Mok TSK, 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. Lancet. 2019;393(10183):1819–30.
8. Herbst RS, Giaccone G, de Marinis F, Reinmuth N, Vergnenegre A, Barrios CH, et al. Atezolizumab for First-Line treatment of PD-L1-Selected patients with NSCLC. N Engl J Med. 2020;383(14):1328–39.
9. Felip E, Altorki N, Zhou C, Csőszi T, Vynnychenko I, Goloborodko O, et al. Adjuvant Atezolizumab after adjuvant chemotherapy in resected stage IB-IIIA non-small-cell lung cancer (IMpower010): a randomised, multicentre, open-label, phase 3 trial. Lancet. 2021;398(10308):1344–57.
10. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic Non-Small-Cell lung Cancer. N Engl J Med. 2018;378(22):2078–92.
11. Wang J, Lu S, Yu X, Hu Y, Sun Y, Wang Z, et al. Tislelizumab plus chemotherapy vs chemotherapy alone as First-line treatment for advanced squamous Non-Small-Cell lung cancer: A phase 3 randomized clinical trial. JAMA Oncol. 2021;7(5):709–17.
12. Zhou C, Chen G, Huang Y, Zhou J, Lin L, Feng J, et al. Camrelizumab plus carboplatin and pemetrexed versus chemotherapy alone in chemotherapy-naive patients with advanced non-squamous non-small-cell lung cancer (CameL): a randomised, open-label, multicentre, phase 3 trial. Lancet Respir Med. 2021;9(3):305–14.
13. Wu X, Huang Y, Zhao Q, Wang L, Song X, Li Y, et al. PD-L1 expression correlation with metabolic parameters of FDG PET/CT and clinicopathological characteristics in non-small cell lung cancer. EJNMMI Res. 2020;10(1):51.
14. Li Y, Xie F, Xiong Q, Lei H, Feng P. Machine learning for lymph node metastasis prediction of in patients with gastric cancer: A systematic review and meta-analysis. Front Oncol. 2022;12:946038.
15. Bedrikovetski S, Dudi-Venkata NN, Kroon HM, Seow W, Vather R, Carneiro G, et al. Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis. BMC Cancer. 2021;21(1):1058.
16. Santer M, Kloppenburg M, Gottfried TM, Runge A, Schmutzhard J, Vorbach SM et al. Current applications of artificial intelligence to classify cervical lymph nodes in patients with head and neck squamous cell Carcinoma-A systematic review. Cancers (Basel). 2022;14(21):5397.
17. Chen J, Chen A, Yang S, Liu J, Xie C, Jiang H. Accuracy of machine learning in preoperative identification of genetic mutation status in lung cancer: A systematic review and meta-analysis. Radiother Oncol. 2024;196:110325.
18. Didier AJ, Nigro A, Noori Z, Omballi MA, Pappada SM, Hamouda DM. Application of machine learning for lung cancer survival prognostication-A systematic review and meta-analysis. Front Artif Intell. 2024;7:1365777.
19. de la Pinta C, Barrios-Campo N, Sevillano D. Radiomics in lung cancer for oncologists. J Clin Transl Res. 2020;6(4):127–34.
20. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.
21. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess the risk of Bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–8.
22. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res. 2019;28(9):2768–86.
23. Meißner AK, Gutsche R, Galldiks N, Kocher M, Jünger ST, Eich ML, et al. Radiomics for the non-invasive prediction of PD-L1 expression in patients with brain metastases secondary to non-small cell lung cancer. J Neurooncol. 2023;163(3):597–605.
24. Wu J, Liu C, Liu X, Sun W, Li L, Gao N, et al. Artificial intelligence-assisted system for precision diagnosis of PD-L1 expression in non-small cell lung cancer. Mod Pathol. 2022;35(3):403–11.
25. Choi S, Cho SI, Ma M, Park S, Pereira S, Aum BJ, et al. Artificial intelligence-powered programmed death ligand 1 analyser reduces interobserver variation in tumour proportion score for non-small cell lung cancer with better prediction of immunotherapy response. Eur J Cancer. 2022;170:17–26.
26. Wang X, Chen P, Ding G, Xing Y, Tang R, Peng C, et al. Dual-scale categorization based deep learning to evaluate programmed cell death ligand 1 expression in non-small cell lung cancer. Med (Baltim). 2021;100(20):e25994.
27. Mu W, Jiang L, Shi Y, Tunali I, Gray JE, Katsoulakis E, et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. J Immunother Cancer. 2021;9(6):e002118.
28. Jiang Z, Dong Y, Yang L, Lv Y, Dong S, Yuan S, et al. CT-Based Hand-crafted radiomic signatures can predict PD-L1 expression levels in Non-small cell lung cancer: a Two-Center study. J Digit Imaging. 2021;34(5):1073–85.
29. Zhao X, Zhao Y, Zhang J, Zhang Z, Liu L, Zhao X. Predicting PD-L1 expression status in patients with non-small cell lung cancer using [(18)F]FDG PET/CT radiomics. EJNMMI Res. 2023;13(1):4.
30. Wang YB, He X, Song X, Li M, Zhu D, Zhang F, et al. The radiomic biomarker in non-small cell lung cancer: (18)F-FDG PET/CT characterisation of programmed death-ligand 1 status. Clin Radiol. 2023;78(10):e732–40.
31. Liu PM, Feng B, Shi JF, Feng HJ, Hu ZJ, Chen YH, et al. A deep-learning model using enhanced chest CT images to predict PD-L1 expression in non-small-cell lung cancer patients. Clin Radiol. 2023;78(10):e689–97.
32. Wang C, Ma J, Shao J, Zhang S, Liu Z, Yu Y, et al. Predicting EGFR and PD-L1 status in NSCLC patients using multitask AI system based on CT images. Front Immunol. 2022;13:813072.
33. Wang C, Ma J, Shao J, Zhang S, Li J, Yan J, et al. Non-Invasive measurement using deep learning algorithm based on Multi-Source features

fusion to predict PD-L1 expression and survival in NSCLC. Front Immunol. 2022;13:828560.
34. Shiinoki T, Fujimoto K, Kawazoe Y, Yuasa Y, Kajima M, Manabe Y, et al. Predicting programmed death-ligand 1 expression level in non-small cell lung cancer using a combination of peritumoral and intratumoral radiomic features on computed tomography. Biomed Phys Eng Express. 2022;8(2):025008.
35. Lim CH, Koh YW, Hyun SH, Lee SJ. A machine learning approach using PET/CT-based radiomics for prediction of PD-L1 expression in Non-small cell lung Cancer. Anticancer Res. 2022;42(12):5875–84.
36. Wang C, Xu X, Shao J, Zhou K, Zhao K, He Y, et al. Deep learning to predict EGFR mutation and PD-L1 expression status in Non-Small-Cell lung Cancer on computed tomography images. J Oncol. 2021;2021:5499385.
37. Hondelink LM, Hüyük M, Postmus PE, Smit V, Blom S, von der Thüsen JH, et al. Development and validation of a supervised deep learning algorithm for automated whole-slide programmed death-ligand 1 tumour proportion score assessment in non-small cell lung cancer. Histopathology. 2022;80(4):635–47.
38. Wen Q, Yang Z, Dai H, Feng A, Li Q. Radiomics study for predicting the expression of PD-L1 and tumor mutation burden in Non-Small cell lung Cancer based on CT images and clinicopathological features. Front Oncol. 2021;11:620246.
39. Tian P, He B, Mu W, Liu K, Liu L, Zeng H, et al. Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images. Theranostics. 2021;11(5):2098–107.
40. Bracci S, Dolciami M, Trobiani C, Izzo A, Pernazza A, D'Amati G, et al. Quantitative CT texture analysis in predicting PD-L1 expression in locally advanced or metastatic NSCLC patients. Radiol Med. 2021;126(11):1425–33.
41. Zhu Y, Liu YL, Feng Y, Yang XY, Zhang J, Chang DD, et al. A CT-derived deep neural network predicts for programmed death ligand-1 expression status in advanced lung adenocarcinomas. Ann Transl Med. 2020;8(15):930.
42. Yoon J, Suh YJ, Han K, Cho H, Lee HJ, Hur J, et al. Utility of CT radiomics for prediction of PD-L1 expression in advanced lung adenocarcinomas. Thorac Cancer. 2020;11(4):993–1004.
43. Sha L, Osinski BL, Ho IY, Tan TL, Willis C, Weiss H, et al. Multi-Field-of-View deep learning model predicts nonsmall cell lung Cancer programmed Death-Ligand 1 status from Whole-Slide hematoxylin and Eosin images. J Pathol Inf. 2019;10:24.
44. Hashimoto K, Murakami Y, Omura K, Takahashi H, Suzuki R, Yoshioka Y, et al. Prediction of tumor PD-L1 expression in resectable Non-Small cell lung Cancer by machine learning models based on clinical and radiological features: performance comparison with preoperative biopsy. Clin Lung Cancer. 2024;25(1):e26–ee346.
45. Zhang R, Hohenforst-Schmidt W, Steppert C, Sziklavari Z, Schmidkonz C, Atzinger A, et al. Standardized 18F-FDG PET/CT radiomic features provide information on PD-L1 expression status in treatment-naïve patients with non-small cell lung cancer. Nuklearmedizin. 2022;61(5):385–93.
46. Fu Y, Zhang H, Xue P, Ren M, Xiao T, Zhang Z, et al. Qualitative analysis of PD-L1 expression in non-small-cell lung cancer based on chest CT radiomics. Biomed Signal Process Control. 2023;84:104815.
47. Sun Z, Hu S, Ge Y, Wang J, Duan S, Song J, et al. Radiomics study for predicting the expression of PD-L1 in non-small cell lung cancer based on CT images and clinicopathologic features. J Xray Sci Technol. 2020;28(3):449–59.
48. Shao J, Ma J, Zhang S, Li J, Dai H, Liang S, et al. Radiogenomic system for Non-Invasive identification of multiple actionable mutations and PD-L1 expression in Non-Small cell lung Cancer based on CT images. Cancers (Basel). 2022;14:19.
49. Cheng G, Zhang F, Xing Y, Hu X, Zhang H, Chen S, et al. Artificial Intelligence-Assisted score analysis for predicting the expression of the immunotherapy biomarker PD-L1 in lung Cancer. Front Immunol. 2022;13:893198.
50. Liu J, Zheng Q, Mu X, Zuo Y, Xu B, Jin Y, et al. Automated tumor proportion score analysis for PD-L1 (22C3) expression in lung squamous cell carcinoma. Sci Rep. 2021;11(1):15907.
51. Li J, Ge S, Sang S, Hu C, Deng S. Evaluation of PD-L1 expression level in patients with Non-Small cell lung Cancer by (18)F-FDG PET/CT radiomics and clinicopathological characteristics. Front Oncol. 2021;11:789014.
52. 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. Acad Radiol. 2020;27(2):171–9.
53. Seol HY, Kim YS, Kim SJ. Predictive value of 18F-fluorodeoxyglucose positron emission tomography/computed tomography for PD-L1 expression in non-small cell lung cancer: A systematic review and meta-analysis. Thorac Cancer. 2020;11(11):3260–8.
54. Chen Q, Zhang L, Mo X, You J, Chen L, Fang J, et al. Current status and quality of radiomic studies for predicting immunotherapy response and outcome in patients with non-small cell lung cancer: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2021;49(1):345–60.
55. Goldberg SB, Schalper KA, Gettinger SN, Mahajan A, Herbst RS, Chiang AC, et al. Pembrolizumab for management of patients with NSCLC and brain metastases: long-term results and biomarker analysis from a non-randomised, open-label, phase 2 trial. Lancet Oncol. 2020;21(5):655–63.
56. Socinski MA, Jotte RM, Cappuzzo F, Orlandi F, Stroyakovskiy D, Nogami N, et al. Atezolizumab for First-Line treatment of metastatic nonsquamous NSCLC. N Engl J Med. 2018;378(24):2288–301.

Publisher's note