scientific reports

Check for updates

OPEN

Predictive value of machine learning for radiation pneumonitis and checkpoint inhibitor pneumonitis in lung cancer patients: a systematic review and meta-analysis

Shenghan Wang[1,2], Kaiyue Wang[1,2] & Jiangnan Lin[1,2]✉

Some studies have developed machine learning (ML) models for the prediction of pneumonitis following immunotherapy and radiotherapy for patients with lung cancer (LC). However, the prediction accuracy of these models remains a topic of debate. Thus, this study aims to summarize the advantages of ML methods in the early prediction of radiation pneumonitis (RP) and checkpoint inhibitor pneumonitis (CIP) in LC patients. PubMed, Cochrane, Embase, and Web of Science were searched up to March 23, 2025. The Prediction Model Risk of Bias Assessment Tool (PROBAST) was utilized to explore the risk of bias (RoB) in the included studies. A subgroup analysis was conducted based on variables including radiomics, dosiomics, and clinical characteristics. Fifty-six studies comprising 12,803 LC patients were included. Of these, 43 studies focused on the early prediction of RP, 11 studies on CIP, and 2 studies on differentiating RP and CIP. The meta-analysis revealed that the c-index of dosiomics-based models, radiomics-based models, and models based on radiomics and clinical characteristics for predicting RP was 0.82 (95% CI: 0.76–0.87), 0.80 (95% CI: 0.71–0.89), and 0.90 (95% CI: 0.86–0.94), respectively. In the prediction of CIP, the c-index for the clinical characteristics model was 0.83 (95% CI: 0.81–0.85), while the integrated radiomics and clinical characteristics model achieved a c-index of 0.86 (95% CI: 0.80–0.92). The ML-based models exhibit strong performance for predicting RP and CIP. Models that integrate dosiomics and radiomics demonstrate superior predictive performance for RP. In addition, hybrid models combining radiomics with clinical features provide excellent predictive value for CIP.

Keywords Machine learning, Radiation pneumonitis, Checkpoint inhibitor pneumonitis, Radiomics

Abbreviations

AI	Artificial intelligence
LC	Lung cancer
NSCLC	Non-small cell lung cancer
ML	Machine learning
RP	Radiation pneumonitis
CIP	Checkpoint inhibitor pneumonitis
SEN	Sensitivity
SPE	Specificity
TS	Training set
VS	Validation set

[1]Department of Radiology, The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), No. 54 Youdian Rd, Hangzhou 310006, Zhejiang, China. [2]First School of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou 310053, Zhejiang, China. ✉email: linjiangnan@zcmu.edu.cn

Lung cancer (LC) remains the most prevalent cancer and the leading cause of cancer-related deaths worldwide, accounting for 18.0% of all cancer-related deaths[1]. Despite significant advances in its diagnosis and treatment, the five-year survival rate for patients with LC remains as low as 22%[1–3]. Non-small cell lung cancer (NSCLC) accounts for 85% of all LC cases and is the most commonly diagnosed type[4].

For LC patients, the primary treatment strategies include surgical resection, chemotherapy, targeted therapy and radiotherapy[5]. Currently, radiotherapy plays a significant role in treating LC, contributing to improved survival outcomes of LC patients[6]. In parallel, researchers have discovered that immunotherapies, particularly immune checkpoint inhibitors (ICIs), have advanced rapidly, providing prolonged survival benefits to NSCLC patients[7]. Recent results also support the potential synergistic effect between radiotherapy and ICI immunotherapy, thus improving patient outcomes[6].

Although radiotherapy and immunotherapy significantly improve survival outcomes, they are known to induce treatment-related toxicities. Radiation pneumonitis (RP) and checkpoint inhibitor pneumonitis (CIP) are clinically significant complications associated with an increased risk of death in affected patients[8,9]. At present, there are currently no effective methods to predict these two toxicities[10,11]. Even though dosiomics is primarily used to predict RP in the early stage[12], its predictive accuracy remains limited. This may be due to the individualization of radiation doses and the inability of dose-volume histograms (DVH) to accurately account for spatial dose distribution or organ structure[13,14].

In recent years, there has been growing interest in machine learning (ML) as it has the potential to develop new prognostic and diagnostic models for LC patients. Zheng et al. investigated radiomics-based models for predicting LC staging[15]. Another review by Kothari G et al. examined the diagnostic accuracy of deep learning models for NSCLC patients undergoing radiotherapy[16]. Building on this foundation, some researchers developed models for the early prediction of RP and CIP in LC patients utilizing ML techniques[17–19]. However, the significant heterogeneity in the selection of ML algorithms and the variability in input features across studies[20–22] have led to inconsistent predictive performance and limited comparability of findings. This inconsistency highlights the need for further exploration of ML methods in predicting RP and CIP in LC patients.

Consequently, this study seeks to evaluate the potential advantages of ML techniques in predicting RP and CIP in LC patients. Specifically, the study will assess the effectiveness of various ML methods based on different modeling variables. Through this approach, the study aims to provide evidence-based insights to support the integration of ML into clinical practice for predicting RP and CIP in LC patients.

Material and methods

Study registration

This research was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) and was registered prospectively with Prospero (CRD 42024521281).

Eligibility criteria

Inclusion criteria

Studies were eligible if they

(1) focused on LC patients who received radiotherapy or immunotherapy.
(2) developed a ML model for predicting RP or CIP in these patients.
(3) were cohort studies, case-control studies, or cross-sectional studies.

Exclusion criteria

Studies were ineligible if they

(1) were meta-analyses, reviews, guidelines, and expert opinions.
(2) focused solely on identifying univariate predictors without developing multivariate predictive models or machine learning-based algorithms for risk stratification.
(3) lacked outcome measures for assessing the accuracy of prediction models, including the following: ROC, c-statistic, c-index, sensitivity (SEN), specificity (SPE), accuracy, recall, precision, confusion matrix, diagnostic fourfold (DF) table, F1 score, or calibration curve.

Data sources and search strategy

We conducted a comprehensive search of the PubMed, Cochrane, Embase, and Web of Science databases as of March 23, 2025. We utilized subject headings and text keywords in this process, with no restrictions on geographical location or publication date (Supplementary Table S1).

Study selection and data extraction

All collected articles were managed by EndNote. After duplicate removal, the remaining articles were screened for further exclusion by reviewing their titles and abstracts. The full texts of potentially eligible articles were downloaded and read to select eligible studies. Before data extraction, we established a spreadsheet, which included the following: title, the first author, publication year, publication country, study type, patient source, tumor type, treatment background, prediction events, diagnostic standard of prediction events, number of prediction events, total cases, the number of prediction cases in the training set (TS), total cases in the TS, validation set (VS) generation method, overfitting method, prediction cases in the VS, total cases in the VS, management of missing values, variable screening/characteristic selection methods, model types, and modeling variables.

The literature screening and data retrieval were conducted independently by two researchers (WSH and WKY). A cross-check was subsequently performed to ensure accuracy. Any disagreements were resolved through discussion with a third researcher (LJN).

Risk of bias assessment

For the included studies, we assessed the risk of bias (RoB) using the Prediction Model Risk of Bias Assessment Tool (PROBAST). PROBAST involves the following four dimensions: subjects, predictors, outcomes, and analysis[23]. Each dimension included a series of questions designed to reflect overall bias risk and applicability. The four dimensions consisted of 2, 3, 6, and 9 questions, respectively. Each question was answered by "yes/probably yes (low risk of bias)," "no/probably not (high risk of bias)," or "no information (unclear)".

The methodological quality of radiomics studies was assessed using the Radiomics Quality Score (RQS)[24]. This evaluation encompasses 16 key components that include data selection, medical imaging, feature extraction, exploratory analysis, and modeling. Each component contributes to the final score, and the total score ranges from − 8 to 36 points. The ROB of all enrolled studies was assessed by three reviewers (WSH, WKY, and LJN).

Outcomes

The primary outcome measure is the c-index, which is used to evaluate the overall accuracy (OA) of the model. However, when the outcome event is characterized by severely unbalanced data, the c-index may be inadequate to predict the model accuracy of positive outcome events. In such cases, the SEN and SPE of the RP or CIP predictive models will also be evaluated.

The accuracy of predictive models is significantly influenced by the selection and processing of modeling variables. Therefore, for the models incorporated in the studies that utilize clinical characteristics, we summarize the specific clinical characteristics accounted for in each model and their frequency of use across all models.

Synthesis methods

For studies that did not report the standard error or 95% confidence interval for the c-index, we estimated these values using the method described by Debray et al.[25]. A random-effects model was preferred when conducting the meta-analysis of the c-index.

Additionally, the meta-analysis of SEN and SPE on the basis of the DF tables was implemented with a bivariate mixed-effects model. However, since many included studies did not provide these tables, the SEN, SPE, precision, and the number of cases or SEN and SPE originated from the best Youden's index and the number of cases were used in calculating the DF table. Interstudy heterogeneity was assessed using Cochran's Q test[26], with significant heterogeneity defined as a p-value < 0.05 or I^2 >50%. The meta-analysis for this study was conducted using R4.2.0 (R development Core Team, Vienna, http://www.R-project.org).

Results

Study selection

The literature retrieval details are illustrated in Fig. 1. 1,322 studies were initially identified through database searches. After excluding 566 duplicates, 756 studies were further screened based on titles and abstracts. Ultimately, our systematic review included 56 studies[8,17–22,27–75] involving a total of 12,803 eligible patients.

Study characteristics

The 56 included studies were mainly published between 2017 and 2025. Among the included studies, 38 were from China, eight from the United States, six from Japan, two from Korea, one from the United Kingdom and one from Turkey. Almost all the studies (55/56, 98%) were cohort studies[8,17–22,27–30,32–75]. Only one study was a case-control study[31]. There were 14 multi-center studies[18,21,22,31,36–38,41,43,56,57,66,67,70], while the rest were single-center studies[8,17,19,20,27–30,32–35,39,40,42,44–55,58–65,68,69,71–75]. Regarding LC types, 31 studies reported NSCLC[18,22,27–31,38,39,42,43,46–51,53,56–62,66,67,70,72,74,75], while the remaining 25 studies did not specifically subdivide LC types[8,17,19–21,32–37,40,41,44,45,52,54,55,63–65,68,69,71,73]. 43 studies[8,17,20,21,27,36–72,75] included 92 models concerning the early prediction of RP, 11 studies[18,19,22,30–35,73,74] included 34 models concerning the early prediction of CIP, and two studies[28,29] included models differentiating between RP and CIP. In the original studies predicting RP, 9,368 LC patients were included, of whom 1,981 patients developed RP. In contrast, studies focused on predicting CIP included 3,435 LC patients, of whom 655 developed CIP. 27 studies were mainly validated by random sampling[18,19,22,27,28,30–34,37,39,43–45,50–52,57,62,65,67,68,71–74], while 12 were externally validated across different centers[17,18,21,22,40,41,48,56,66,67,70,71]. In the constructed models, 27 studies focused on radiomics[17,19–22,28,29,31,35,37–42,44–46,48,54,61,62,65–67,73,74], while 36 studies explored dosiomics[8,20,21,27,36,37,39,41–45,47–52,54–62,64–72], and four studies also included genomics[38,56,60,68]. The basic characteristics of the included studies are presented in Supplementary Table S2.

Risk of bias assessment

Our study evaluated the risk of bias across all 130 models included in the analysis. Seven of these models exhibited a high risk of bias in subject selection due to their derivation from a single non-nested case-control study[31]. It remained unclear whether the predictive factors used in these seven models were assessed with prior knowledge of the outcomes. Therefore, the risk of bias in this domain remained uncertain. Nevertheless, all 130 models exhibited a low risk of bias in the results domain, as they satisfied the PROBAST criteria for results-related questions. A high bias risk was observed in the statistical analysis, primarily because it was difficult to meet the following requirements: event per variable (EPV) ≥ 20 and a sample size exceeding 100 in the independent VS. Furthermore, the complexity and data weight were insufficiently described. The detailed quality evaluation is presented in Fig. 2.

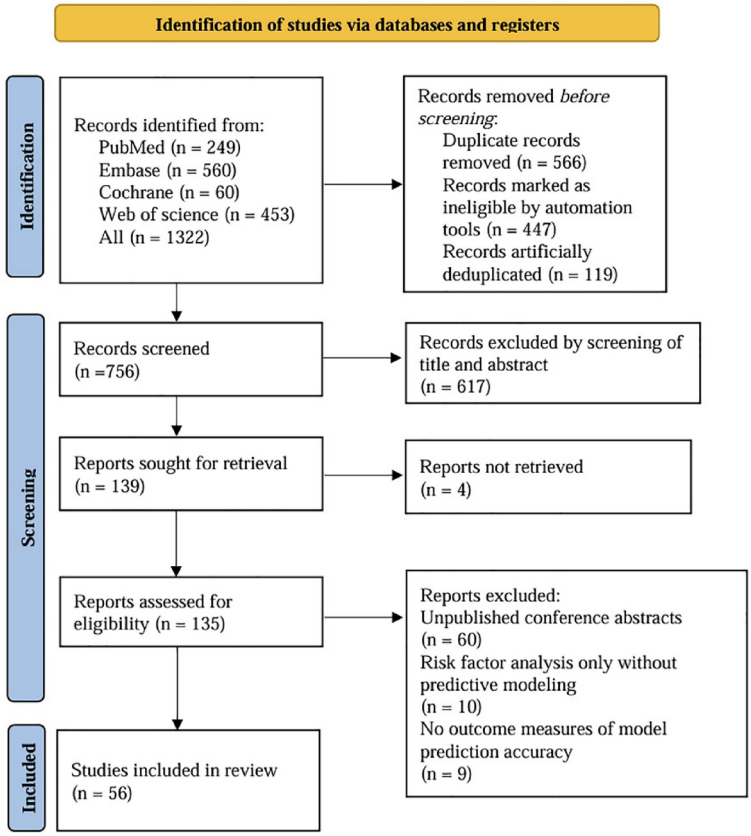

Fig. 1. Basic characteristics of included studies.

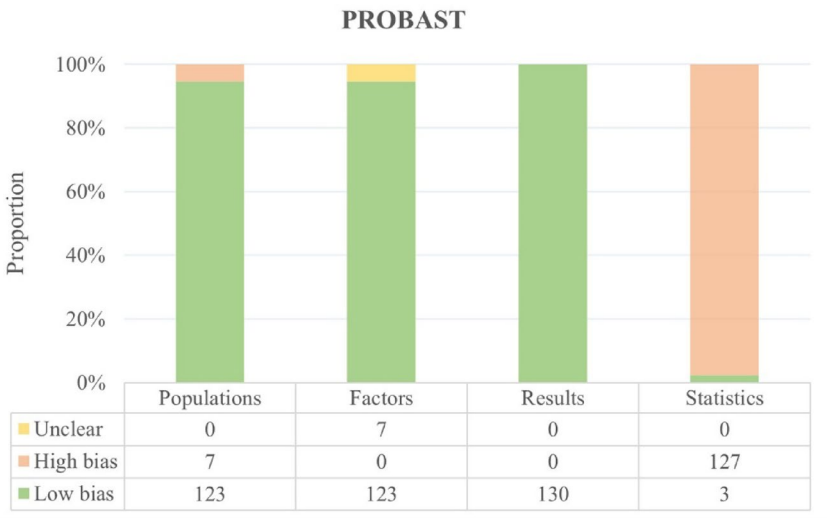

Fig. 2. Risk of bias assessment of included models.

The total RQS and the percentage of RQS for each study are summarized in Table S3. The median RQS for these studies was 12 (range: 9–18). The corresponding percentage of RQS was 33.33% (range: 25-50%). Figure 3 illustrates the percentages of scores in these studies for the 16 components of the RQS.

Meta-analysis

RP prediction

Currently, the models for predicting RP were primarily established based on clinical characteristics, dosiomics, radiomics, or combinations of these factors.

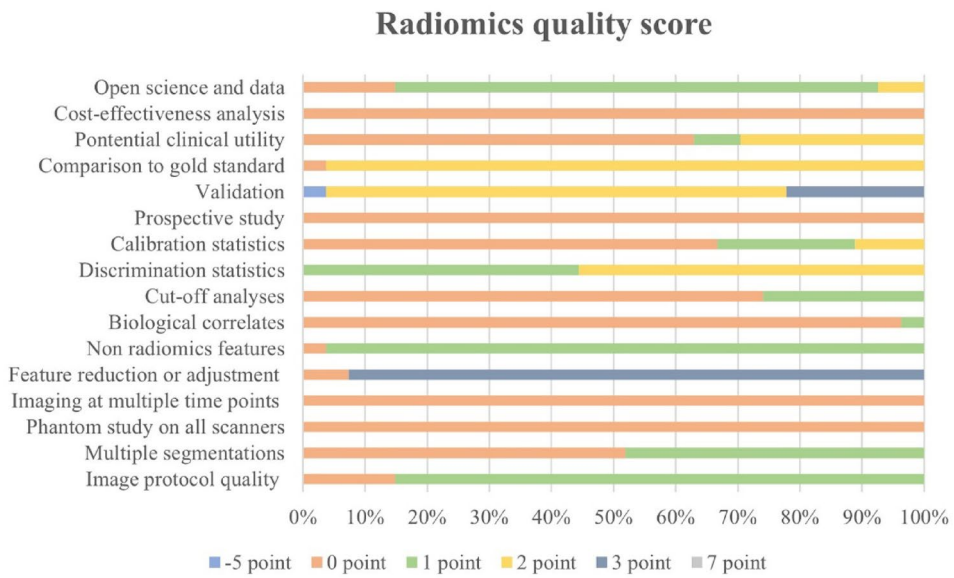

Fig. 3. Percentage stacked bar chart of radiomics quality score.

Synthesized results For the 8,737 patients included in the TS model predicting RP, the pooled c-index was 0.87 (95% CI: 0.85–0.88). Heterogeneity, assessed using the Cochran's Q test, revealed significant variability ($I^2 = 99.4\%$, $p < 0.001$), as shown in Fig. 4. The SEN was 0.83 (0.78–0.87, $I^2 = 80.9\%$), while the SPE was 0.86 (95% CI: 0.81–0.89, $I^2 = 95.8\%$), as illustrated in Fig. 5. In the VS that included 2,106 patients, the pooled c-index was 0.82 (95% CI: 0.79–0.85, $I^2 = 90.7\%$), and the SEN and SPE were 0.81 (95% CI: 0.76–0.85, $I^2 = 38.3\%$) and 0.77 (95% CI: 0.72–0.81, $I^2 = 83.5\%$), respectively (Figs. 6 and 7).

Subgroup analysis In predicting RP, we performed a subgroup analysis based on different modeling variables. In the TS, the dosiomics-based model (1,423 patients) achieved a c-index of 0.85 (95% CI: 0.77–0.93, $I^2 = 99.1\%$), while the model based on clinical characteristics (618 patients) achieved a c-index of 0.80 (95% CI: 0.73–0.86, $I^2 = 46.4\%$). The radiomics-based model (1,017 patients) achieved a c-index of 0.92 (95% CI: 0.88–0.96, $I^2 = 89.8\%$), compared to 0.81 (95% CI: 0.76–0.87, $I^2 = 94.9\%$) for the dosiomics combined with clinical characteristics model (2,680 patients). The radiomics combined with dosiomics model (1,468 patients) showed a c-index of 0.92 (95% CI: 0.86–0.98, $I^2 = 99.8\%$) in Fig. 4. In the VS, c-indices were 0.82 (95% CI: 0.76–0.87, $I^2 = 94.4\%$) for the dosiomics-based model (543 patients), 0.80 (95% CI: 0.71–0.89, $I^2 = 73.3\%$) for the radiomics-based model (462 patients), 0.78 (95% CI: 0.66–0.90, $I^2 = 94.5\%$) for the dosiomics combined with clinical characteristics model (328 patients), and 0.90 (95% CI: 0.86–0.94, $I^2 = 28.2\%$) the radiomics combined with dosiomics model (397 patients), as depicted in Fig. 6. SEN and SPE values are available in Figures S1-S9.

Reporting bias The funnel plot of our TS indicated the presence of publication bias, whereas the funnel plot of the VS was symmetrical (Figures S10-S11).

CIP prediction

The models for predicting CIP were mainly based on clinical characteristics, radiomics, and the combination of radiomics with clinical characteristics.

Synthesized results The pooled c-index for the predictive model in the CIP training set with 5,564 patients was 0.85 (95% CI: 0.83–0.88, $I^2 = 75.6\%$), with SEN and SPE values of 0.80 (95% CI: 0.76–0.82, $I^2 = 9.9\%$) and 0.82 (95% CI: 0.79–0.85, $I^2 = 88.4\%$), respectively (Figure S12-S13). In the VS that included 2,758 patients, the pooled c-index was 0.84 (95% CI: 0.82–0.87, $I^2 = 54.4\%$), while the SEN and SPE were 0.72 (95% CI: 0.68–0.76, $I^2 = 0\%$) and 0.84 (95% CI: 0.80–0.87, $I^2 = 70.1\%$), as shown in Figure S14-15.

Subgroup analysis For CIP prediction, a subgroup analysis was also implemented in accordance with different modeling variables. In the TS, the model based on clinical characteristics (2,409 patients) achieved a c-index of 0.82 (95% CI: 0.80–0.84, $I^2 = 34.7\%$), compared to 0.89 (95% CI: 0.87–0.92, $I^2 = 0\%$) for the radiomics-based model (492 patients) and 0.88 (95% CI: 0.82–0.93, $I^2 = 82.1\%$) for the model based on radiomics combined with clinical characteristics (698 patients), as illustrated in Figure S12. In the VS, the model based on clinical characteristics (1,179 patients) achieved a c-index of 0.83 (95% CI: 0.81–0.85, $I^2 = 4.3\%$), compared to 0.84 (95% CI: 0.81–0.86, $I^2 = 0\%$) for the radiomics-based model (226 patients) and 0.86 (95% CI: 0.80–0.92, $I^2 = 72.9\%$) for the model based on radiomics combined with clinical characteristics (337 patients), as depicted in Figure S14. The values for SEN and SPE are presented in Figures S16-S21.

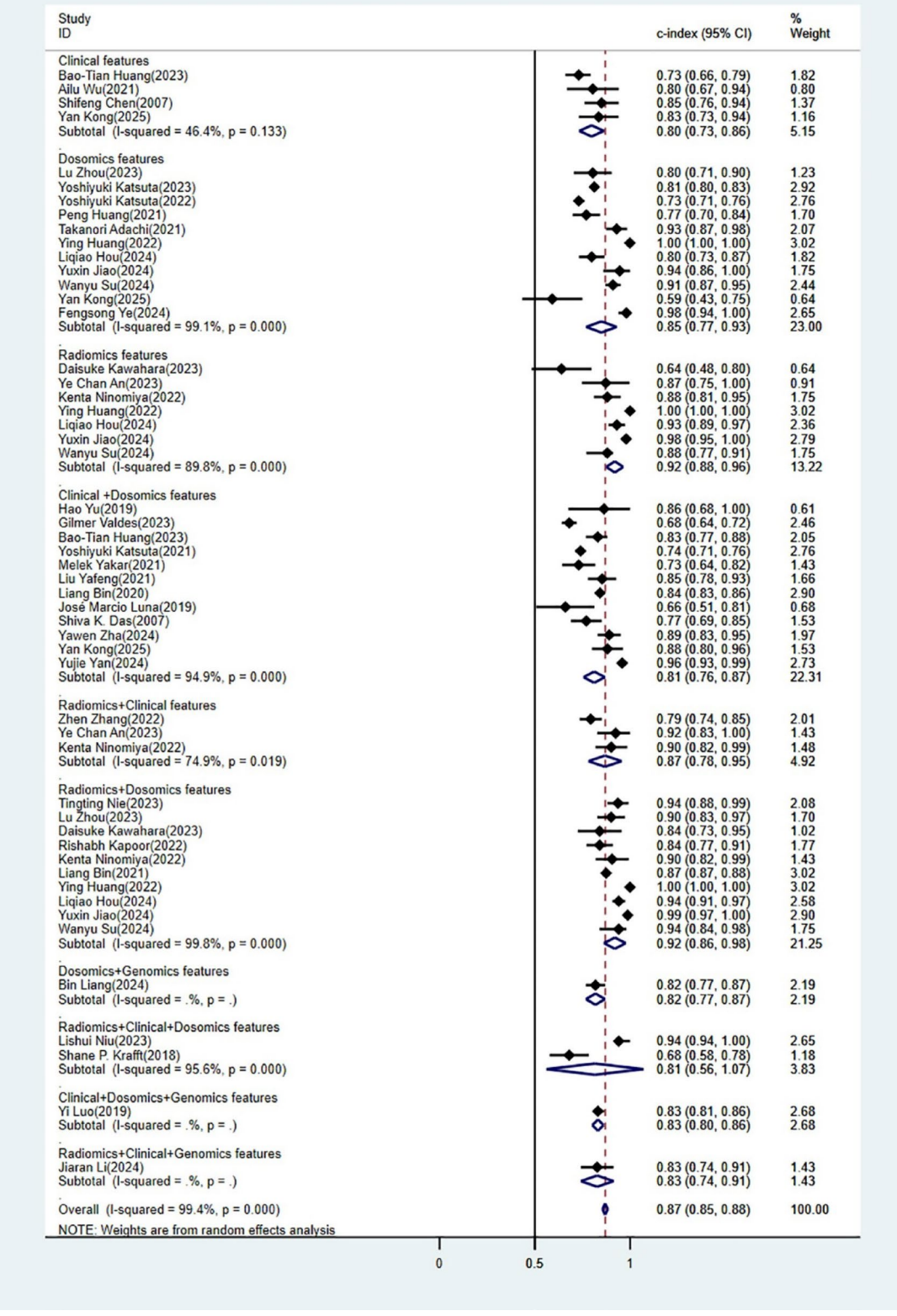

Fig. 4. Forest plot of c-index meta-analysis of ML model for RP prediction in the TS.

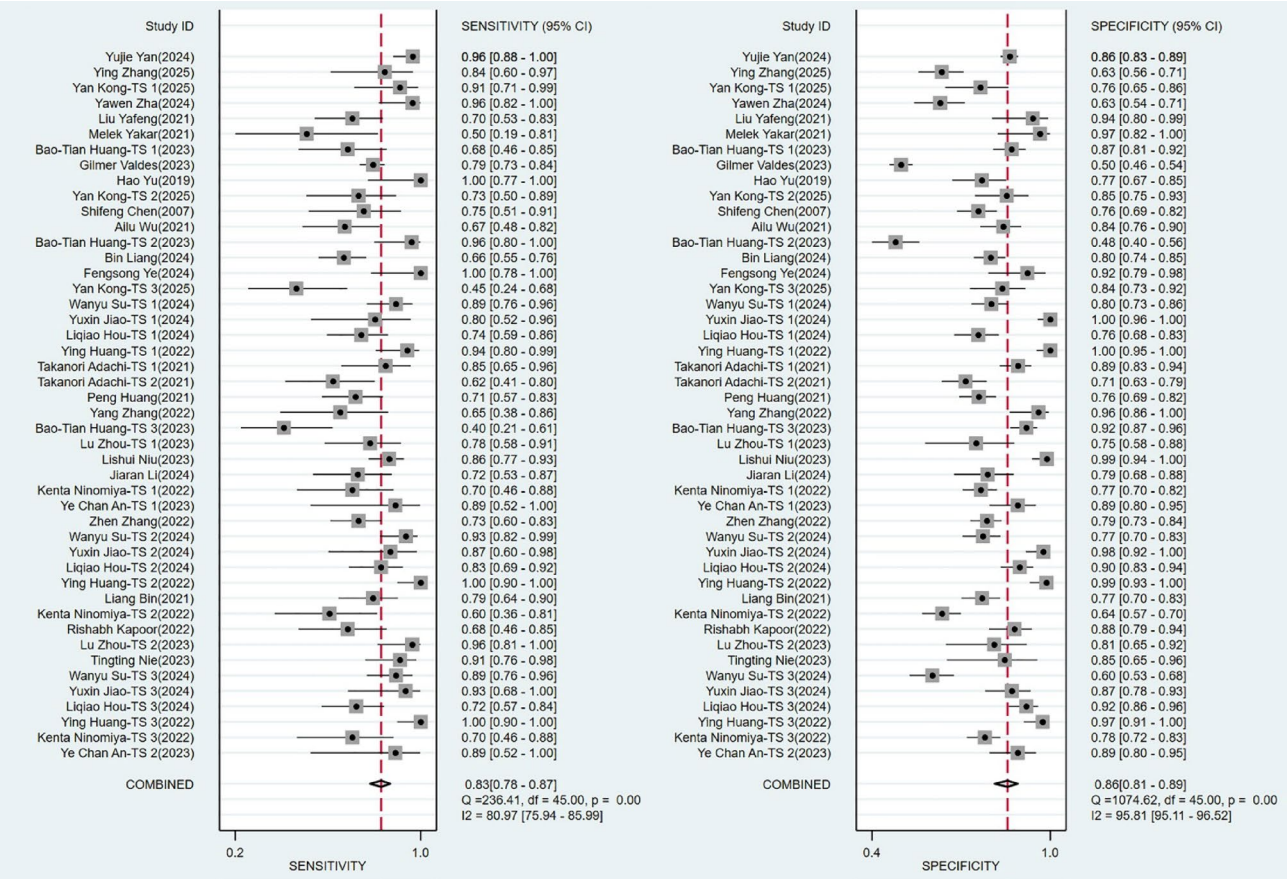

Fig. 5. Forest plot of SEN and SPE meta-analysis of ML model for RP prediction in the TS.

Reporting bias The funnel plot of our training set and validation set did not show any obvious evidence of asymmetry, indicating no publication bias (Figure S22-S23).

CIP and RP identification

Two studies have developed models for distinguishing between RP and CIP[28,29]. However, due to the limited number of studies, we did not conduct a relevant meta-analysis. In these two studies, the AUC for identifying RP and CIP was 0.95 (95% CI: 0.92–0.99) and 0.88 (95% CI: 0.82–0.95), respectively.

Modeling variables

We summarized the common clinical characteristics used to construct the models. Age is frequently utilized as a predictive variable in models for RP, appearing in seven models[8,17,45,50,53,61,69]. Interstitial lung disease (ILD) is the most commonly used clinical feature in models predicting CIP, and it is incorporated into five distinct models[18,22,31,32]. A comprehensive list of all variables and their frequencies can be found in Tables S4 and S5. Figures S24 and S25 visually depict the distribution and frequency of these clinical features.

Discussion

Summary of the main findings

Our analysis reveals that ML-based models demonstrate superior predictive performance for predicting RP and CIP in LC patients. Specifically, integrating dosiomics with radiomics achieved a higher c-index compared to dosiomics alone in RP prediction. Furthermore, incorporating radiomics with clinical characteristics provided better predictive performance than using clinical features alone for predicting CIP.

Comparison with previous other reviews

Our findings align with recent systematic reviews on the application of ML in RP prediction, while also offering several analytical improvements. Chen et al.[76] conducted a meta-analysis of ML models incorporating multimodal features, reporting a pooled AUC of 0.93 for 9 studies ($n = 1,406$), with SEN and SPE of 0.74 and 0.91, respectively. Although the research conducted by Chen et al. demonstrated the potential of multimodal integration, their findings need to be validated due to their small sample size and a lack of detailed analysis on the qualitative advantages for specific feature combinations. In contrast, our systematic review includes 43 studies ($n = 9,368$) for RP prediction, significantly enhancing statistical power and generalizability. Notably, unlike the analysis performed by Chen et al., who broadly categorized features as multimodal, we quantitatively

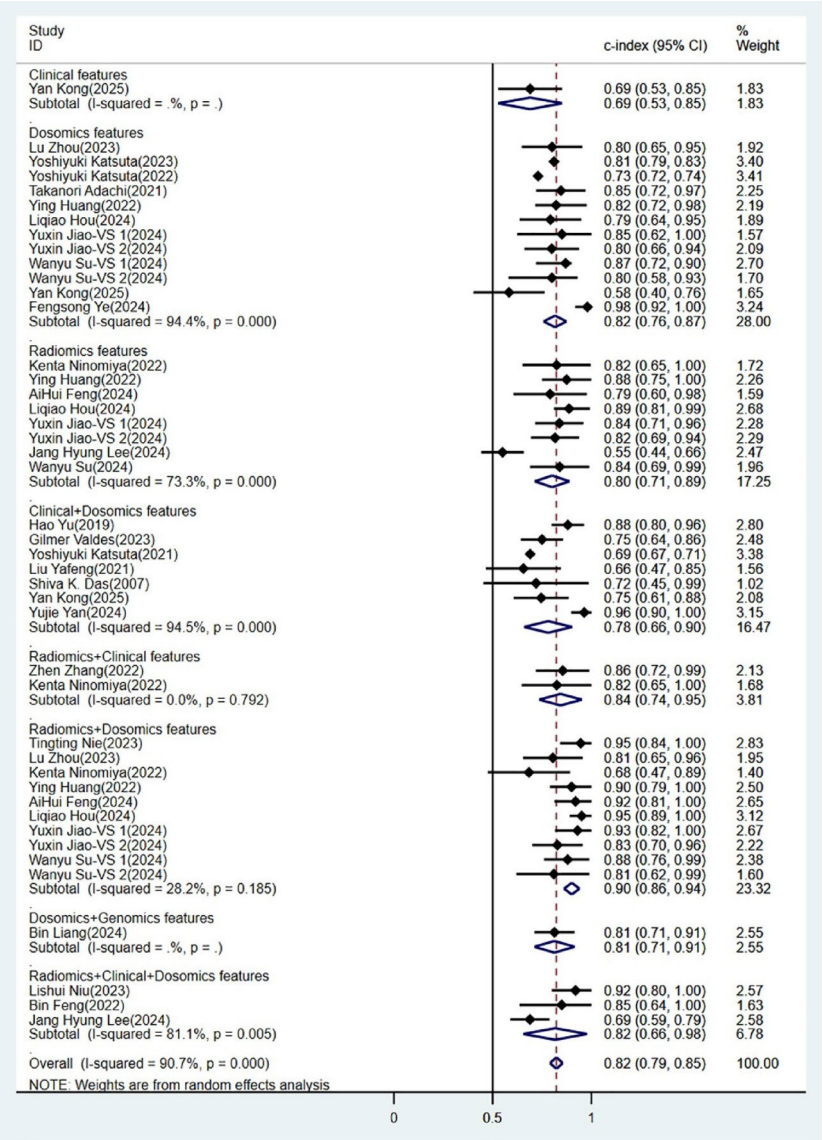

Fig. 6. Forest plot of c-index meta-analysis of ML model for RP prediction in the VS.

assessed the impact of individual modalities and their combinations on predictive performance. For example, our analysis reveals that models based on radiomics plus dosiomics achieved a superior c-index of 0.90 (95% CI: 0.86–0.94) in the validation cohort. This combined model significantly outperforms models based only on dosiomics (c-index = 0.82) or radiomics (c-index = 0.80). These results highlight the synergistic value of multimodal data, a nuance not systematically examined in previous reviews.

In predicting CIP, our findings indicate that a hybrid model combining clinical features and radiomics (c-index = 0.86) outperforms the model that incorporates only clinical features (c-index = 0.83) in terms of predictive performance, further validating the universality of multimodal integration. This result is consistent with research across multiple fields. For instance, various multimodal biomarkers that integrate clinical and imaging data have been shown to outperform unimodal models in terms of prognostic accuracy for patients with metastatic NSCLC receiving immunotherapy[77]. In the stroke field, multimodal ML models similarly demonstrate superior diagnostic accuracy[78]. Furthermore, a scoping review indicates that multimodal fusion can improve the average AUC by 6.4%[79], which aligns with our findings. In our study, multimodal models achieved an 8.8% higher c-index for predicting RP compared to unimodal approaches (0.90 vs. 0.82). Research across multiple fields has confirmed the broad applicability of multimodal integration. This approach helps mitigate the information loss inherent in unimodal analysis and capture complementary biological signals, thereby enabling more precise patient stratification and risk prediction[80,81].

It is important to acknowledge that the high I^2 statistic in our study indicates substantial heterogeneity among the included studies. This methodological challenge is particularly evident in subgroup analysis. The persistent heterogeneity highlights existing barriers to clinical implementation of radiomics-based and dosiomics-based machine learning predictive models. Specifically, our low RQS results highlight the lack of rigor in the radiomics

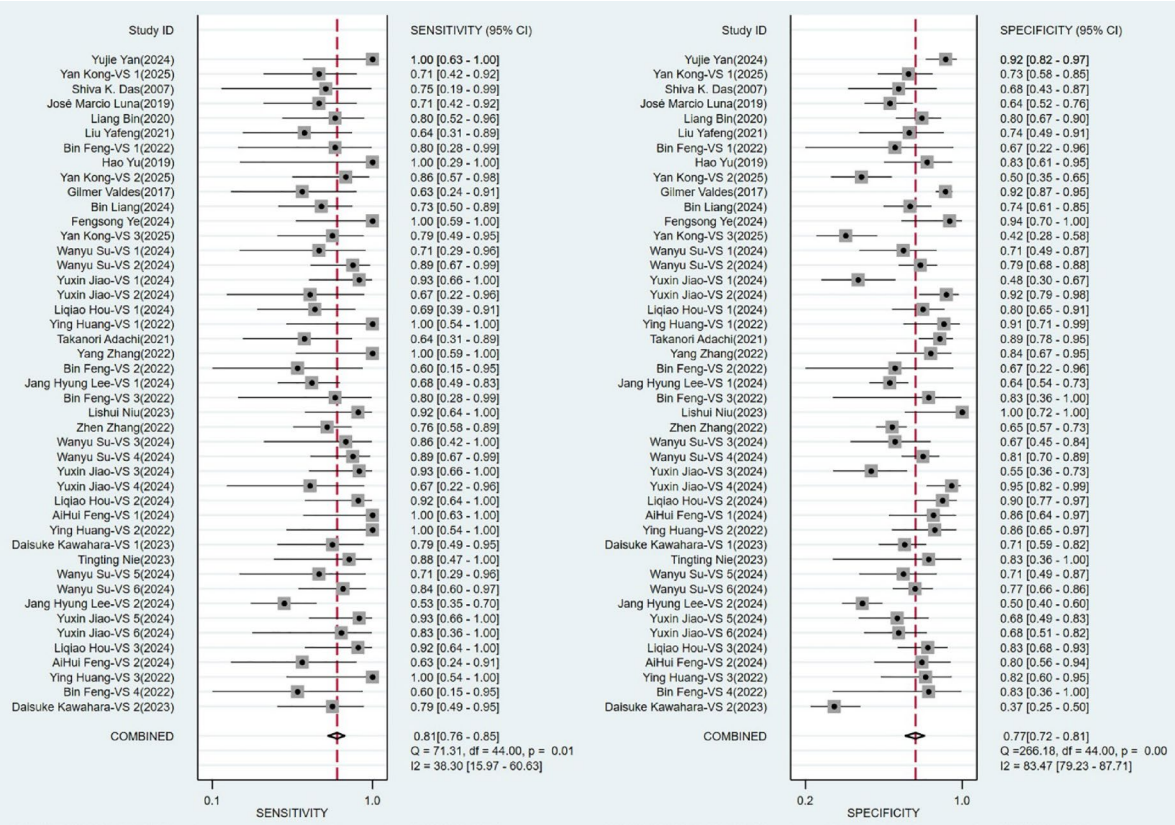

Fig. 7. Forest plot of SEN and SPE meta-analysis of ML model for RP prediction in the VS.

workflow, particularly in the areas of image acquisition and feature extraction, which can lead to discrepancies in the representation of radiomic features. These discrepancies ultimately affect model performance[82–84]. In addition, the inconsistency in radiation dose parameters across studies may contribute to this heterogeneity, as variations in dose metrics can influence the relationship between radiation exposure and RP predictions[13,55,70,72]. Moreover, all studies included in this review employed ML techniques. Summarizing these studies inevitably involves some degree of heterogeneity. Because when training ML models, random parameter initialization can affect model convergence and reproducibility[85]. Despite these challenges, the results of our meta-analysis remain valid, as long as we acknowledge the sources of heterogeneity[85]. By analyzing the potential sources of heterogeneity, we are able to draw broader conclusions and improve predictive models by considering the impacts of these heterogeneous factors. Moving forward, standardizing radiomics procedures and improving model training methods could help mitigate some of this heterogeneity, leading to more robust and reproducible models.

Highlights and limitations of the study

Our study is the first to comprehensively evaluate the predictive efficiency of machine learning for RP and CIP, establishing evidence-based support for the use of artificial intelligence in this domain. However, the current systematic review has several limitations, and the results should be interpreted with caution. Firstly, this review is retrospective, and thus it is impossible to analyze feature importance and evaluate multivariate interactions. The reason is that most original studies do not report standardized variable weights and do not furnish individual participant data. Secondly, due to the inconsistencies in feature selection strategies and validation frameworks, it is difficult to explore synergistic mechanisms between clinical predictors and radiomics or dosiomics. Thirdly, due to disproportionate focus on discrimination metrics over calibration parameters and limited access to raw datasets, it is impossible to ascertain model calibration across studies. Fourthly, significant methodological differences, including inconsistent radiomic feature extraction methodologies, may contribute to uncertainty and heterogeneity observed in the joint model estimates. Fifthly, varying degrees of risk of bias are observed between the included studies. Given the above limitations, it is necessary to develop standardized reporting protocols in radiomics research and the development of predictive models encompassing the disclosure of model hyperparameters, calibration statistics, and reproducible analytical workflows.

Conclusions

Our research demonstrates that ML-based models are effective in predicting RP and CIP. Notably, models based on both dosiomics and radiomics offer superior predictive performance for RP. Additionally, models based on both clinical features and radiomics also exhibit significant predictive advantages for CIP. Standardizing the

radiomics process and the development of predictive models is essential for enhancing the clinical applicability of predictive models for RP and CIP.

Data availability

The original contributions presented in the study are included in the article/Supplementary Material.

Received: 16 January 2025; Accepted: 3 June 2025
Published online: 01 July 2025

References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **72**, 7–33. https://doi.org/10.3322/caac.21708 (2022).
2. Thai, A. A., Solomon, B. J., Sequist, L. V., Gainor, J. F. & Heist, R. S. Lung cancer. *Lancet* **398**, 535–554. https://doi.org/10.1016/s0140-6736(21)00312-3 (2021).
3. Sung, H. et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249. https://doi.org/10.3322/caac.21660 (2021).
4. She, Y. et al. Development and validation of a deep learning model for Non-Small cell lung Cancer survival. *JAMA Netw. Open.* **3**, e205842. https://doi.org/10.1001/jamanetworkopen.2020.5842 (2020).
5. Li, Y., Yan, B. & He, S. Advances and challenges in the treatment of lung cancer. *Biomed. Pharmacother.* **169**, 115891. https://doi.org/10.1016/j.biopha.2023.115891 (2023).
6. Vinod, S. K. & Hau, E. Radiotherapy treatment for lung cancer: current status and future directions. *Respirology* **25**, 61–71. https://doi.org/10.1111/resp.13870 (2020).
7. Reck, M. et al. First-line nivolumab plus ipilimumab with two cycles of chemotherapy versus chemotherapy alone (four cycles) in advanced non-small-cell lung cancer: checkmate 9LA 2-year update. *ESMO Open.* **6**, 100273. https://doi.org/10.1016/j.esmoop.2021.100273 (2021).
8. Huang, B. T., Lin, P. X., Wang, Y. & Luo, L. M. Developing a Prediction Model for Radiation Pneumonitis in Lung Cancer Patients Treated With Stereotactic Body Radiation Therapy Combined With Clinical, Dosimetric Factors, and Laboratory Biomarkers. *Clin Lung Cancer* 24, e323-e331. https://doi.org/10.1016/j.cllc.2023.08.007 (2023).
9. Sears, C. R. et al. Knowledge gaps and research priorities in immune checkpoint Inhibitor-related pneumonitis. An official American thoracic society research statement. *Am. J. Respir Crit. Care Med.* **200**, e31–e43. https://doi.org/10.1164/rccm.201906-1202ST (2019).
10. Hanania, A. N., Mainwaring, W., Ghebre, Y. T., Hanania, N. A. & Ludwig, M. Radiation-Induced lung injury: assessment and management. *Chest* **156**, 150–162. https://doi.org/10.1016/j.chest.2019.03.033 (2019).
11. Kalisz, K. R., Ramaiya, N. H., Laukamp, K. R. & Gupta, A. Immune checkpoint inhibitor Therapy-related pneumonitis: patterns and management. *Radiographics* **39**, 1923–1937. https://doi.org/10.1148/rg.2019190036 (2019).
12. Palma, D. A. et al. Predicting radiation pneumonitis after chemoradiation therapy for lung cancer: an international individual patient data meta-analysis. *Int. J. Radiat. Oncol. Biol. Phys.* **85**, 444–450. https://doi.org/10.1016/j.ijrobp.2012.04.043 (2013).
13. Bourbonne, V. et al. Radiomics analysis of 3D dose distributions to predict toxicity of radiotherapy for lung cancer. *Radiother Oncol.* **155**, 144–150. https://doi.org/10.1016/j.radonc.2020.10.040 (2021).
14. van Baardwijk, A. et al. Mature results of an individualized radiation dose prescription study based on normal tissue constraints in stages I to III non-small-cell lung cancer. *J. Clin. Oncol.* **28**, 1380–1386. https://doi.org/10.1200/jco.2009.24.7221 (2010).
15. Zheng, X. et al. Diagnostic accuracy of deep learning and radiomics in lung Cancer staging: A systematic review and Meta-Analysis. *Front. Public Health.* **10**, 938113. https://doi.org/10.3389/fpubh.2022.938113 (2022).
16. Kothari, G. et al. A systematic review and meta-analysis of the prognostic value of radiomics based models in non-small cell lung cancer treated with curative radiotherapy. *Radiother. Oncol.* **155**, 188–203. https://doi.org/10.1016/j.radonc.2020.10.023 (2021).
17. Zhang, Z. et al. Computed tomography and radiation dose images-based deep-learning model for predicting radiation pneumonitis in lung cancer patients after radiation therapy. *Radiother. Oncol.* **182**, 109581. https://doi.org/10.1016/j.radonc.2023.109581 (2023).
18. Jia, X. et al. Predicting checkpoint inhibitors pneumonitis in non-small cell lung cancer using a dynamic online hypertension nomogram. *Lung Cancer.* **170**, 74–84. https://doi.org/10.1016/j.lungcan.2022.06.001 (2022).
19. Cheng, M. et al. Deep learning for predicting the risk of immune checkpoint inhibitor-related pneumonitis in lung cancer. *Clin. Radiol.* **78**, e377–e385. https://doi.org/10.1016/j.crad.2022.12.013 (2023).
20. Lee, J. H. et al. Deep-Learning model prediction of radiation pneumonitis using pretreatment chest computed tomography and clinical factors. *Technol. Cancer Res. Treat.* **23**, 15330338241254060. https://doi.org/10.1177/15330338241254060 (2024).
21. Su, W. et al. Multi-omics deep learning for radiation pneumonitis prediction in lung cancer patients underwent volumetric modulated Arc therapy. *Comput. Methods Progr. Biomed.* **254**, 108295. https://doi.org/10.1016/j.cmpb.2024.108295 (2024).
22. Du, Y. et al. Radiomics biomarkers to predict checkpoint inhibitor pneumonitis in Non-small cell lung Cancer. *Acad. Radiol.* **32**, 1685–1695. https://doi.org/10.1016/j.acra.2024.09.053 (2025).
23. Moons, K. G. M. et al. A tool to assess risk of Bias and applicability of prediction model studies: explanation and elaboration. *Ann. Intern. Med.* **170**, W1–w33. https://doi.org/10.7326/m18-1377 (2019).
24. Lambin, P. et al. Radiomics: the Bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762. https://doi.org/10.1038/nrclinonc.2017.141 (2017).
25. Debray, T. P. et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat. Methods Med. Res.* **28**, 2768–2786. https://doi.org/10.1177/0962280218785504 (2019).
26. Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *Bmj* **327**, 557–560. https://doi.org/10.1136/bmj.327.7414.557 (2003).
27. Yu, H. et al. Machine learning to build and validate a model for radiation pneumonitis prediction in patients with Non-Small cell lung Cancer. *Clin. Cancer Res.* **25**, 4343–4350. https://doi.org/10.1158/1078-0432.Ccr-18-1084 (2019).
28. Qiu, Q., Xing, L., Wang, Y., Feng, A. & Wen, Q. Development and validation of a radiomics nomogram using computed tomography for differentiating immune checkpoint Inhibitor-Related pneumonitis from radiation pneumonitis for patients with Non-Small cell lung Cancer. *Front. Immunol.* **13**, 870842. https://doi.org/10.3389/fimmu.2022.870842 (2022).
29. Peiliang Wang, M. D., Li, Y., Zhao, M. M. M. & Yu, M. M. J. Feifei teng, M. D. Distinguishing immune checkpoint inhibitor-related pneumonitis from radiation pneumonitis by CT radiomics features in non-small cell lung cancer. *Int. Immunopharmacol.* **128**, 111489. https://doi.org/10.1016/j.intimp.2024.111489 (2024).
30. Zhang, Y. et al. A nomogram model for predicting the risk of checkpoint inhibitor-related pneumonitis for patients with advanced non-small-cell lung cancer. *Cancer Med.* **12**, 15998–16010. https://doi.org/10.1002/cam4.6244 (2023).
31. Zhang, G. Y. et al. Development and validation of a machine Learning-Based model using CT radiomics for predicting immune checkpoint Inhibitor-related pneumonitis in patients with NSCLC receiving Anti-PD1 immunotherapy: A multicenter retrospective casecontrol study. *Acad. Radiol.* **31**, 2128–2143. https://doi.org/10.1016/j.acra.2023.10.039 (2024).

32. Jia, X. et al. Comprehensive nomogram models for predicting checkpoint inhibitor pneumonitis. *Am. J. Cancer Res.* **13**, 2681–2701 (2023).
33. Hong, B., Chen, R., Zheng, C., Liu, M. & Yang, J. Development and validation of a nomogram for predicting immune-related pneumonitis after sintilimab treatment. *Cancer Med.* **13**, e6708. https://doi.org/10.1002/cam4.6708 (2024).
34. Gong, L. et al. Identification and prediction of immune checkpoint inhibitors-related pneumonitis by machine learning. *Front. Immunol.* **14**, 1138489. https://doi.org/10.3389/fimmu.2023.1138489 (2023).
35. Tan, P. et al. Deep learning predicts immune checkpoint inhibitor-related pneumonitis from pretreatment computed tomography images. *Front. Physiol.* **13**, 978222. https://doi.org/10.3389/fphys.2022.978222 (2022).
36. Valdes, G. et al. Predicting the effect of proton beam therapy technology on pulmonary toxicities for patients with locally advanced lung Cancer enrolled in the proton collaborative group prospective clinical trial. *Int. J. Radiat. Oncol. Biol. Phys.* **119**, 66–77. https://doi.org/10.1016/j.ijrobp.2023.11.026 (2024).
37. Nie, T. et al. Integration of dosimetric parameters, clinical factors, and radiomics to predict symptomatic radiation pneumonitis in lung cancer patients undergoing combined immunotherapy and radiotherapy. *Radiother. Oncol.* **190**, 110047. https://doi.org/10.1016/j.radonc.2023.110047 (2024).
38. Li, J. et al. Novel model integrating computed tomography-based image markers with genetic markers for discriminating radiation pneumonitis in patients with unresectable stage III non-small cell lung cancer receiving radiotherapy: a retrospective multi-center radiogenomics study. *BMC Cancer.* **24**, 78. https://doi.org/10.1186/s12885-023-11809-y (2024).
39. Zhou, L. et al. Machine Learning-Based Multiomics Prediction Model for Radiation Pneumonitis. *J Oncol* 5328927. https://doi.org/10.1155/2023/5328927 (2023).
40. Zhang, Z. et al. Radiomics and dosiomics signature from whole lung predicts radiation pneumonitis: A model development study with prospective external validation and Decision-curve analysis. *Int. J. Radiat. Oncol. Biol. Phys.* **115**, 746–758. https://doi.org/10.1016/j.ijrobp.2022.08.047 (2023).
41. Niu, L. et al. A multiomics approach-based prediction of radiation pneumonia in lung cancer patients: impact on survival outcome. *J. Cancer Res. Clin. Oncol.* **149**, 8923–8934. https://doi.org/10.1007/s00432-023-04827-7 (2023).
42. Kawahara, D., Imano, N., Nishioka, R. & Nagata, Y. Image masking using convolutional networks improves performance classification of radiation pneumonitis for non-small cell lung cancer. *Phys. Eng. Sci. Med.* **46**, 767–772. https://doi.org/10.1007/s13246-023-01249-0 (2023).
43. Katsuta, Y. et al. Radiation pneumonitis prediction model with integrating multiple dose-function features on 4DCT ventilation images. *Phys. Med.* **105**, 102505. https://doi.org/10.1016/j.ejmp.2022.11.009 (2023).
44. Kapoor, R., Sleeman, W., Palta, J. & Weiss, E. 3D deep Convolution neural network for radiation pneumonitis prediction following stereotactic body radiotherapy. *J. Appl. Clin. Med. Phys.* **24**, e13875. https://doi.org/10.1002/acm2.13875 (2023).
45. Feng, B. et al. Pseudo-siamese network combined with dosimetric and clinical factors, radiomics features, CT images and 3D dose distribution for the prediction of radiation pneumonitis: A feasibility study. *Clin. Transl. Radiat. Oncol.* **38**, 188–194. https://doi.org/10.1016/j.ctro.2022.11.011 (2023).
46. An, Y. C. et al. Quantification of diffuse parenchymal lung disease in non-small cell lung cancer patients with definitive concurrent chemoradiation therapy for predicting radiation pneumonitis. *Thorac. Cancer.* **14**, 3530–3539. https://doi.org/10.1111/1759-7714.15156 (2023).
47. Zhang, Y. et al. Development and validation of a random forest model for predicting radiation pneumonitis in lung cancer patients receiving moderately hypofractionated radiotherapy: a retrospective cohort study. *Ann. Transl. Med.* **10**, 1264. https://doi.org/10.21037/atm-22-3049 (2022).
48. Ninomiya, K., Arimura, H., Yoshitake, T., Hirose, T. A. & Shioyama, Y. Synergistic combination of a topologically invariant imaging signature and a biomarker for the accurate prediction of symptomatic radiation pneumonitis before stereotactic ablative radiotherapy for lung cancer: A retrospective analysis. *PLoS One.* **17**, e0263292. https://doi.org/10.1371/journal.pone.0263292 (2022).
49. Katsuta, Y. et al. Feasibility of differential Dose-Volume histogram features in multivariate prediction model for radiation pneumonitis occurrence. *Diagnostics* 12. https://doi.org/10.3390/diagnostics12061354 (2022).
50. Katsuta, Y. et al. Prediction of radiation pneumonitis with machine learning using 4D-CT based dose-function features. *J. Radiat. Res.* **63**, 71–79. https://doi.org/10.1093/jrr/rrab097 (2022).
51. Yakar, M., Etiz, D., Metintas, M., Ak, G. & Celik, O. Prediction of radiation pneumonitis with machine learning in stage III lung cancer: A pilot study. *Technol. Cancer Res. Treat.* **20**, 15330338211016373. https://doi.org/10.1177/15330338211016373 (2021).
52. Yafeng, L. et al. Construction and verification of a radiation pneumonia prediction model based on multiple parameters. *Cancer Control.* **28**, 10732748211026671. https://doi.org/10.1177/10732748211026671 (2021).
53. Wu, A., Zhou, Z., Song, Y., Liang, S. & Li, F. Application of a radiation pneumonitis prediction model in patients with locally advanced lung squamous cell cancer. *Ann. Palliat. Med.* **10**, 4409–4417. https://doi.org/10.21037/apm-21-459 (2021).
54. Bin, L. et al. A deep learning-based dual-omics prediction model for radiation pneumonitis. *Med. Phys.* **48**, 6247–6256. https://doi.org/10.1002/mp.15079 (2021).
55. Huang, P. et al. Predicting radiation pneumonitis with fuzzy clustering neural network using 4DCT ventilation image based dosimetric parameters. *Quant. Imaging Med. Surg.* **11**, 4731–4741. https://doi.org/10.21037/qims-20-1095 (2021).
56. Cui, S. & Haken, T. El naqa, I. Integrating multiomics information in deep learning architectures for joint actuarial outcome prediction in Non-Small cell lung Cancer patients after radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **110**, 893–904. https://doi.org/10.1016/j.ijrobp.2021.01.042 (2021).
57. Adachi, T. et al. Multi-institutional dose-segmented dosiomic analysis for predicting radiation pneumonitis after lung stereotactic body radiation therapy. *Med. Phys.* **48**, 1781–1791. https://doi.org/10.1002/mp.14769 (2021).
58. Liang, B. et al. Prediction of radiation pneumonitis with dose distribution: A convolutional neural network (CNN) based model. *Front. Oncol.* **9**, 1500. https://doi.org/10.3389/fonc.2019.01500 (2019).
59. Luna, J. M. et al. Predicting radiation pneumonitis in locally advanced stage II-III non-small cell lung cancer using machine learning. *Radiother. Oncol.* **133**, 106–112. https://doi.org/10.1016/j.radonc.2019.01.003 (2019).
60. Cui, S., Luo, Y., Tseng, H. H. & Haken, T. El naqa, I. Combining handcrafted features with latent variables in machine learning for prediction of radiation-induced lung damage. *Med. Phys.* **46**, 2497–2511. https://doi.org/10.1002/mp.13497 (2019).
61. Krafft, S. P. et al. The utility of quantitative CT radiomics features for improved prediction of radiation pneumonitis. *Med. Phys.* **45**, 5317–5324. https://doi.org/10.1002/mp.13150 (2018).
62. Huang, Y. et al. Radiation pneumonitis prediction after stereotactic body radiation therapy based on 3D dose distribution: dosiomics and/or deep learning-based radiomics features. *Radiat. Oncol.* **17**, 188. https://doi.org/10.1186/s13014-022-02154-8 (2022).
63. Chen, S. et al. A neural network model to predict lung radiation-induced pneumonitis. *Med. Phys.* **34**, 3420–3427. https://doi.org/10.1118/1.2759601 (2007).
64. Das, S. K. et al. Predicting lung radiotherapy-induced pneumonitis using a model combining parametric Lyman probit with nonparametric decision trees. *Int. J. Radiat. Oncol. Biol. Phys.* **68**, 1212–1221. https://doi.org/10.1016/j.ijrobp.2007.03.064 (2007).
65. Feng, A. et al. Improvement of prediction performance for radiation pneumonitis by using 3-Dimensional dosiomic features. *Clin. Lung Cancer.* **25**, e173. e180.e172 (2024).

66. Hou, L. et al. CT-based different regions of interest radiomics analysis for acute radiation pneumonitis in patients with locally advanced NSCLC after chemoradiotherapy. *Clin. Transl. Radiat. Oncol.* **48**, 100828. https://doi.org/10.1016/j.ctro.2024.100828 (2024).
67. Jiao, Y. et al. Development and validation of a lung biological equivalent dose-based multiregional radiomic model for predicting symptomatic radiation pneumonitis after SBRT in lung cancer patients. *Front. Oncol.* **14**, 1489217. https://doi.org/10.3389/fonc.2024.1489217 (2024).
68. Liang, B. et al. Synergizing the interaction of single nucleotide polymorphisms with dosiomics features to build a dual-omics model for the prediction of radiation pneumonitis. *Radiother. Oncol.* **196**, 110261. https://doi.org/10.1016/j.radonc.2024.110261 (2024).
69. Zha, Y. et al. A dynamic nomogram predicting symptomatic pneumonia in patients with lung cancer receiving thoracic radiation. *BMC Pulm. Med.* **24**, 99. https://doi.org/10.1186/s12890-024-02899-w (2024).
70. Kong, Y. et al. Enhancing the prediction of symptomatic radiation pneumonitis for locally advanced non-small-cell lung cancer by combining 3D deep learning-derived imaging features with dose-volume metrics: a two-center study. *Strahlenther. Onkol.* **201**, 274–282. https://doi.org/10.1007/s00066-024-02221-x (2025).
71. Yan, Y. et al. Clinical predictors of severe radiation pneumonitis in patients undergoing thoracic radiotherapy for lung cancer. *Transl. Lung Cancer Res.* **13**, 1069–1083. https://doi.org/10.21037/tlcr-24-328 (2024).
72. Ye, F. et al. Predicting radiation pneumonitis in lung cancer: a EUD-based machine learning approach for volumetric modulated Arc therapy patients. *Front. Oncol.* **14**, 1343170. https://doi.org/10.3389/fonc.2024.1343170 (2024).
73. Wang, N., Zhao, Z., Duan, Z. & Xie, F. Predicting immune checkpoint Inhibitor-Related pneumonitis via computed tomography and Whole-Lung analysis deep learning. *Curr. Med. Imaging*. **20**, e15734056314192. https://doi.org/10.2174/0115734056314192241002075034 (2024).
74. Wang, X. et al. Quantification of preexisting lung ground glass opacities on CT for predicting checkpoint inhibitor pneumonitis in advanced non-small cell lung cancer patients. *BMC Cancer*. **24**, 269. https://doi.org/10.1186/s12885-024-12008-z (2024).
75. Valdes, G., Solberg, T. D., Heskel, M., Ungar, L. & Simone, C. B. 2 Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy. *Phys. Med. Biol.* **61**, 6105–6120. https://doi.org/10.1088/0031-9155/61/16/6105 (2016).
76. Chen, Z. et al. Predicting radiation pneumonitis in lung cancer using machine learning and multimodal features: a systematic review and meta-analysis of diagnostic accuracy. *BMC Cancer*. **24**, 1355. https://doi.org/10.1186/s12885-024-13098-5 (2024).
77. Captier, N. et al. Integration of clinical, pathological, radiological, and transcriptomic data improves prediction for first-line immunotherapy outcome in metastatic non-small cell lung cancer. *Nat. Commun.* **16**, 614. https://doi.org/10.1038/s41467-025-55847-5 (2025).
78. Shurrab, S. et al. Multimodal machine learning for stroke prognosis and diagnosis: A systematic review. *IEEE J. Biomed. Health Inf.* **28**, 6958–6973. https://doi.org/10.1109/jbhi.2024.3448238 (2024).
79. Kline, A. et al. Multimodal machine learning in precision health: A scoping review. *NPJ Digit. Med.* **5**, 171. https://doi.org/10.1038/s41746-022-00712-8 (2022).
80. Llinas-Bertran, A., Butjosa-Espín, M., Barberi, V. & Seoane, J. A. Multimodal data integration in early-stage breast cancer. *Breast* **80**, 103892. https://doi.org/10.1016/j.breast.2025.103892 (2025).
81. Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J. & Shah, S. P. Harnessing multimodal data integration to advance precision oncology. *Nat. Rev. Cancer*. **22**, 114–126. https://doi.org/10.1038/s41568-021-00408-3 (2022).
82. Lambin, P. et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer*. **48**, 441–446. https://doi.org/10.1016/j.ejca.2011.11.036 (2012).
83. Sheen, H. et al. Radiomics-based hybrid model for predicting radiation pneumonitis: A systematic review and meta-analysis. *Phys. Med.* **123**, 103414. https://doi.org/10.1016/j.ejmp.2024.103414 (2024).
84. Chen, Q. 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*. **49**, 345–360. https://doi.org/10.1007/s00259-021-05509-7 (2021).
85. Smith, L. A. et al. Machine learning and deep learning predictive models for long-term prognosis in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Lancet Digit. Health*. **5**, e872–e881. https://doi.org/10.1016/s2589-7500(23)00177-2 (2023).

Author contributions

All authors contributed to the study conception and design. Writing - original draft preparation: SW; Writing - review and editing: [SW, KW, JL]; Conceptualization: [SW, KW, JL]; Methodology: [SW, KW, JL]; Formal analysis and investigation: [SW, KW, JL]; Funding acquisition: JL; Resources: [SW, KW, JL]; Supervision: [SW, KW, JL], and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

The study was supported by Research Project of Zhejiang Chinese Medicine University Affiliated Hospital [2022FSYYZY04] and National College Students Innovation and Entrepreneurship Training Program [202410344072X].

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-05505-z.

Correspondence and requests for materials should be addressed to J.L.