Paradigm Academic Press
Journal of Innovations in Medical Research
ISSN 2788-7022
FEB. 2025 VOL.4, NO.1

The Value of Radiomics in Preoperative Identification of Histological Subtypes and Ki-67 Levels in Lung Cancer: A Systematic Review and Meta-Analysis

Ziya Zhao[1]

[1] MD, Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China

Correspondence: Ziya Zhao, MD, Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China.

doi:10.56397/JIMR/2025.02.06

Abstract

Background: Recently, early non-invasive identification of Ki-67 levels and histological subtypes in non-small cell lung cancer remains a significant obstacle. With the application of radiomics in cancer diagnosis and treatment, several researchers have investigated the accuracy of radiomics for non-invasive detection of Ki-67 levels and histological subtypes in lung cancer. Nonetheless, there is a dearth of systematic evidence. Hence, we reviewed the value and accuracy of radiomics for early non-invasive identification of Ki-67 levels and histological subtypes in lung cancer. ***Methods:*** PubMed, Cochrane, Embase, and Web of Science were comprehensively searched till 10 December 2023, using the Radiomics Risk of Bias Assessment Tool. Subgroup analyses of modeling variables were performed. ***Results:*** Thirty-three papers were finally included, with 13 for identifying adenocarcinoma and squamous, and 12 for identifying different pathotypes. In the validation set of the dichotomous task, the meta-analysis results for discriminating high Ki-67 levels yielded 0.77 c-index (95% CI: 0.74-0.79), 0.75 sensitivity (95% CI: 0.70-0.79), and 0.74 specificity (95% CI: 0.70-0.78). The validation set analysis in discriminating lung adenocarcinoma from squamous cell carcinoma yielded 0.78 c-index (95% CI: 0.76-0.80), 0.78 sensitivity (95% CI: 0.70-0.84), and 0.79 specificity (95% CI: 0.73-0.85). ***Conclusion:*** Radiomics is tool for non-invasively identifying high Ki-67 levels in lung cancer and identifying different subtypes. However, this is based on limited evidence with a high risk of bias in a subset of studies where radiomics has been performed. Therefore, more studies with larger samples are required to validate the results and develop intelligent readable tools.

Keywords: lung cancer, radiomics, histological subtypes, Ki-67, systematic review

1. Introduction

Lung cancer is the predominant contributor to cancer morbidity and mortality, with about 2.5 million new cases and over 1.8 million deaths worldwide in 2024, and the most common histologic subtype is lung adenocarcinoma (LADC) (F. Bray, et al., 2024). Currently, surgical interventions remain a cornerstone, including total lung, lobectomy, segmental resection, and wedge resection. The advent of minimally invasive techniques has popularized sub-lobar resection for small early-stage lung cancers (J.S. Donington, et al., 2018; M.K. Kamel et al., 2022; L. Crinò et al., 2010). In addition to surgical treatment, alternative modalities like chemotherapy, radiotherapy, immunotherapy, and targeted therapy are also available for the comprehensive management of lung cancer (N. Duma, R. Santana-Davila & J.R. Molina, 2019), which comprises diverse subtypes necessitating individualized treatment (W.D. Travis et al., 2016).

The 2021 WHO histologic classification (M.S. Tsao et al., 2022) has clarified lung cancer into small cell lung

cancer and non-small cell lung cancer (NSCLC), of which NSCLC accounts for nearly 80-85%, including LADC, lung squamous cell carcinomas (LSCC), and other histologic subtypes. The treatment modalities for different subtypes are diverse, and the treatment for NSCLC relies on molecular markers (M. Jamal-Hanjani et al., 2015). Ki-67 is the extensively utilized protein marker for assessing tumor cell proliferation (Z. Li et al., 2021; A. Warth et al., 2014), and its elevation connects with metastasis and unfavorable outcomes (Z. Li et al., 2021; D.M. Wei et al., 2018). Thus, it is valuable to efficiently identify Ki-67 and histological subtypes before surgery. Currently, post-surgical pathology and biopsy assessment remain the gold standard for identifying tumor histological subtypes and Ki-67 levels. Nevertheless, these methods are invasive and some tumors are not available for biopsy or are resistant to invasive testing. Therefore, exploring an efficient non-invasive approach to precisely predict lung cancer pathologic subtypes and Ki-67 expression is of profound clinical significance.

Currently, radiomics has garnered significant interest and investigation in the diagnosis and efficacy evaluation of lung cancer. This approach relies on automated high-throughput extraction of quantitative features from numerous medical images to quantitatively analyze medical image data and provide doctors with more and better clinical information (P. Lambin et al., 2012). Several studies have utilized CT image-based radiomic profiling to preoperatively forecast Ki-67 levels in lung cancer (Q. Fu et al., 2021; F. Liu et al., 2023; J. Bao et al., 2022). Radiomics holds immense promise in categorizing lung cancer histologic subtypes based on CT images (A. Brunetti et al., 2022; F. Song et al., 2023; H.H. Li et al., 2021). Nevertheless, there remains a dearth of comprehensive evidence regarding its efficacy in radiomics. Hence, this paper was conducted to address this gap.

2. Material and Methods

2.1 Study Registration

Our study followed the PRISMA guidelines and was registered on Prospero.

2.2 Eligibility Criteria

The articles were enrolled if

(1) The study subjects were confirmed NSCLC patients.

(2) In the original study, machine learning (ML) models covering radiomics identification of Ki-67 and histological subtypes were constructed in full.

(3) The types of studies covered case-control studies, cohort studies, and cross-sectional studies.

(4) The article was reported in English.

The articles were excluded for

(1) Meta, review, guideline, or expert opinion;

(2) Only differential factor analysis and no complete ML model;

(3) Lacking endpoint indicators of predictive accuracy of ML models (accuracy, c-index, Roc, c-statistic, sensitivity, specificity, confusion matrix, recall, precision, F1 score, diagnostic four-cell table, Calibration curve);

(4) sample size <20 cases;

(5) Only segmentation of images.

3. Search Strategy

PubMed, Cochrane, Embase, and Web of Science were comprehensively retrieved as of 10 December 2023 using subject plus free word searching, with no restriction on region, language, or years of experience (A comprehensive search strategy is presented in Table S1).

3.1 Article Selection and Data Extraction

All retrieved publications were imported into Endnote. The titles or abstracts were read after excluding duplicates. The original literature that initially matched was downloaded for full-text assessment. Before data extraction, we created a standard spreadsheet containing title, first author, publication year, country, tumor type, study type, patient source, diagnostic purpose, tumor stage, radiomics source, recording of complete image protocols, number of imaging investigators, pre-tests with different imaging parameters, repeated measurement trials at different times, imaging region of interest segmentation software, number of all ending time cases, number of all cases, number of ending event cases in training sets, number of cases in training sets, generation method of validation sets, number of ending event cases in validation sets, number of cases in validation sets, method of variable screening, model type, construction of radiomics scores, assessment of overfitting, public availability of code and data, and model rating metrics. The above article selection and data extraction were undertaken independently by two investigators, and a third investigator decided on any disputes after

examination.

3.2 Assessment of Study Quality

The methodological quality and risk of bias of enrolled literature were appraised by 2 investigators using the radiomics quality score (RQS), and cross-checked upon completion. Any disputes were tackled via discussion with a third investigator.

3.3 Outcomes

Our outcome metrics encompassed c-index, sensitivity, and specificity to reflect the power of radiomics in identifying histological subtypes and Ki-67 in lung cancer.

3.4 Synthesis Methods

A meta-analysis of c-index was performed by which the overall accuracy of ML models was assessed. For some original studies, when 95% confidence intervals (CI) and standard errors were missed, the standard errors were estimated according to Debray TP et al. (2019) Given the different variables and parameters across ML models, random-effects models were utilized in our meta-analysis of c-index, while bivariate mixed-effects models were utilized for meta-analyses of sensitivity and specificity based on the diagnostic quadrangle table. If the diagnostic quadrangle table was not reported, it was calculated using the following two ways: 1) using the combination of the number of cases with sensitivity, specificity, and precision (Precision); 2) based on the best Youden index to extract the sensitivity and specificity, which were then combined with the number of cases for calculation. The meta-analysis was implemented in R4.2.0.

4. Results

4.1 Study Selection

8,557 documents were acquired from various databases, with 2,590 duplicates excluded. 1,490 duplicates were automatically flagged by the software, while 2,441 duplicates were flagged manually. The remaining 5,967 documents were assessed for full texts, with 5,919 excluded for irrelevant topics. Among the remaining 48 preliminary matches, articles with missing data, lacking relevant outcome indicators, and articles related to PET-CT studies were excluded. Thirty-three radiomics articles were finally included (Q. Fu et al., 2021; F. Liu et al., 2023; J. Bao et al., 2022; A. Brunetti et al., 2022; F. Song et al., 2023; H.H. Li et al., 2021; H. Sun et al., 2023; A. Haga et al., 2018; J. Lin et al., 2023; C. Alvarez-Jimenez et al., 2020; B. Dunn et al., 2023; H. Liu et al., 2019; R. Patil et al., 2016; X. Tang et al., 2022; P. Marentakis et al., 2021; M. Zhu et al., 2022; D.D. Yu et al., 2017; M. Yang et al., 2023; F.C. Yang et al., 2021; J. Yan et al., 2022; F. Song et al., 2023; G. Pasini et al., 2023; J. Liu et al., 2019; H. Liu et al., 2021; E. Linning et al., 2019; Z. Khodabakhshi et al., 2021; R. Han et al., 2020; Y.X. Guo et al., 2021; Q.B. Gu et al., 2019; L.N. E et al., 2019; Y. Dong et al., 2022; X.J. Chu et al., 2023; Z.Y. Chen et al., 2022) (Figure 1).

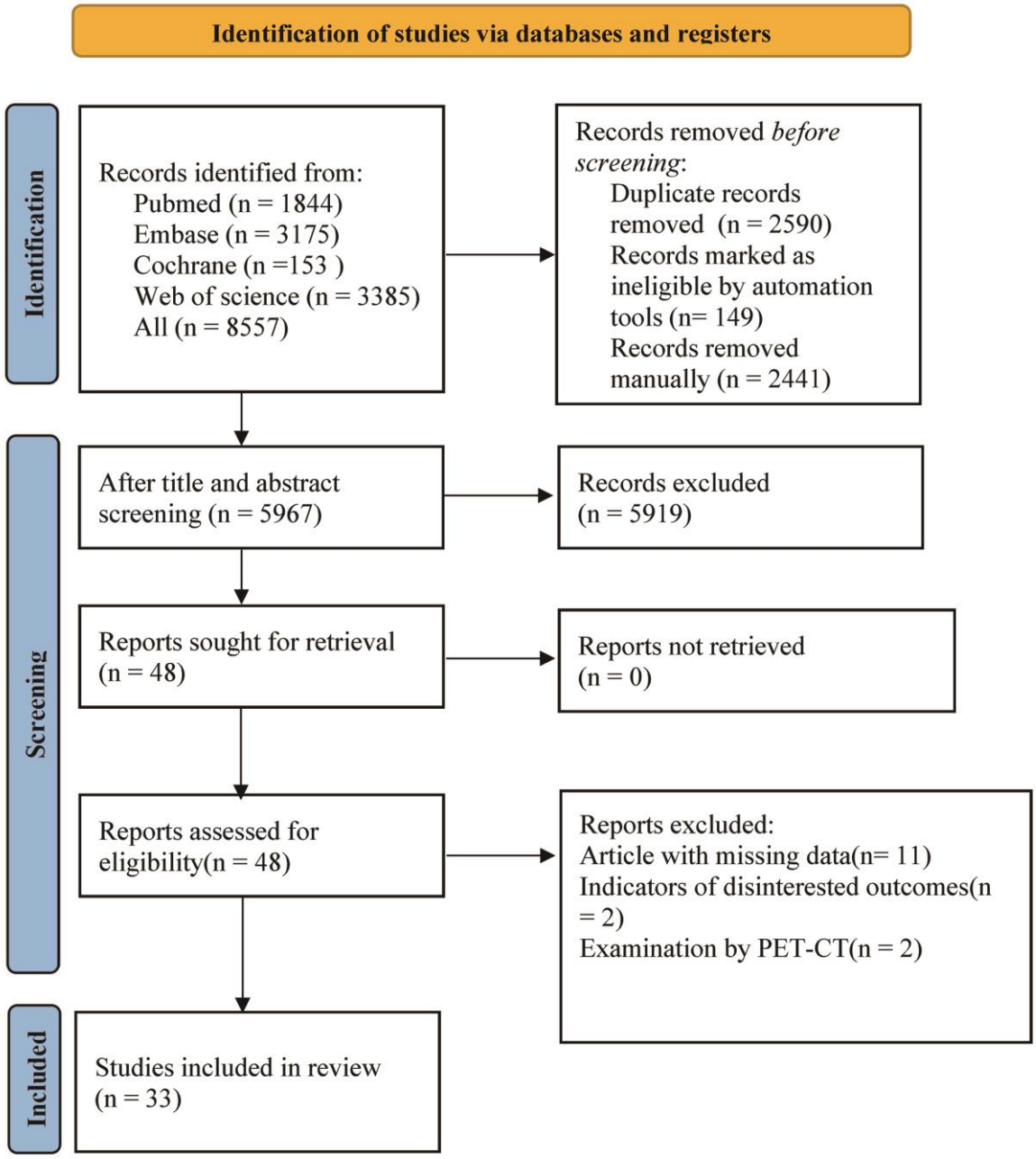


Figure 1. Article screening process

4.2 Study Characteristics

33 papers were included involving 11,037 patients with lung cancer. Of these, 23 articles (F. Liu et al., 2023; F. Song et al., 2023; H. Sun et al., 2023; A. Haga et al., 2018; J. Lin et al., 2023; C. Alvarez-Jimenez et al., 2020; H. Liu et al., 2019; R. Patil et al., 2016; X. Tang et al., 2022; P. Marentakis et al., 2021; D.D. Yu et al., 2017; M. Yang et al., 2023; F.C. Yang et al., 2021; F. Song et al., 2023; G. Pasini et al., 2023; J. Liu et al., 2019; H. Liu et al., 2021; Z. Khodabakhshi et al., 2021; R. Han et al., 2020; Q.B. Gu et al., 2019; Y. Dong et al., 2022; X.J. Chu et al., 2023; Z.Y. Chen et al., 2022) focused on NSCLC, 7 (Q. Fu et al., 2021; A. Brunetti et al., 2022; H.H. Li et al., 2021; B. Dunn et al., 2023; E. Linning et al., 2019; Y.X. Guo et al., 2021; L.N. E et al., 2019) focused on unspecified types of lung cancer, and 3 (J. Bao et al., 2022; T.P. Debray et al., 2019; J. Yan, et al., 2022) focused on LADC. Most were case-control studies. Of the included studies, 5 (F. Liu et al., 2023; A. Brunetti et al., 2022; F.C. Yang et al., 2021; Y. Dong et al., 2022; X.J. Chu et al., 2023) were derived from multicenter databases, 8 (J. Lin et al., 2023; C. Alvarez-Jimenez et al., 2020; B. Dunn et al., 2023; R. Patil et al., 2016; P. Marentakis et al., 2021; F. Song et al., 2023; G. Pasini et al., 2023; J. Liu et al., 2019) were derived from registry databases, 1 (F.

Song et al., 2023) was derived from a multicenter and registry database, and the remaining 19 were all from single-center databases. In terms of diagnostic purposes, 8 (Q. Fu et al., 2021; F. Liu et al., 2023; J. Bao et al., 2022; H. Sun et al., 2023; M. Zhu et al., 2022; J. Yan et al., 2022; Q.B. Gu et al., 2019; Y. Dong et al., 2022) were concerned with Ki-67 expression, while 25 were designed to differentiate pathological subtypes. The main pathotypes included LADC, LSCC, and lung adeno-squamous carcinoma (LASC). Thirteen (A. Brunetti et al., 2022; F. Song et al., 2023; C. Alvarez-Jimenez et al., 2020; H. Liu et al., 2019; X. Tang et al., 2022; P. Marentakis et al., 2021; D.D. Yu et al., 2017; M. Yang et al., 2023; F.C. Yang et al., 2021; F. Song et al., 2023; H. Liu et al., 2021; R. Han et al., 2020; Z.Y. Chen et al., 2022) articles were designed to differentiate between ADC and SCC, with a two-classification primary diagnosis. Additionally, 12 articles (A. Brunetti et al., 2022; A. Haga et al., 2018; J. Lin et al., 2023; B. Dunn et al., 2016; R. Patil et al., 2016; G. Pasini et al., 2023; J. Liu et al., 2019; E. Linning et al., 2019; Z. Khodabakhshi et al., 2021; Y.X. Guo et al., 2021; L.N. E et al., 2019; X.J. Chu et al., 2023) differentiated between ADC, SCC, and other pathotypes. Regarding cancer stages, 13 (J. Bao et al., 2022; A. Brunetti et al., 2022; H. Sun et al., 2023; A. Haga et al., 2018; J. Lin et al., 2023; C. Alvarez-Jimenez et al., 2020; B. Dunn et al., 2023; R. Patil et al., 2016; P. Marentakis et al., 2021; J. Yan et al., 2022; Y.X. Guo et al., 2021; L.N. E et al., 2019; Y. Dong, et al., 2022) focused on stage I-IV tumors, three (J. Bao et al., 2022; A. Haga et al., 2018; J. Yan et al., 2022) focused on early-stage tumors (stages I-II), and 20 papers lacked explicit tumor stage information. Most radiomic data were derived from CT or MRI scans, with most being CT scans. Additionally, two studies (H. Liu et al., 2019; R. Han et al., 2020) employed CE-CT, two (H. Sun et al., 2023; Z.Y. Chen et al.,2022) employed DECT, and one study combined both CT and MRI data.

Of the 11,037 cases included, 4,273 were ADC, 3,142 were SCC, 460 were small-cell lung carcinomas (SCLC), 540 were large-cell carcinomas (LCC), 81 were ASC, and 235 were not otherwise specified (NOS). In the validation method of the model, 17 articles (Q. Fu et al., 2021; F. Liu et al., 2023; J. Bao et al., 2022; H. Sun et al., 2023; A. Haga et al., 2018; J. Lin et al., 2023; X. Tang et al., 2022; P. Marentakis et al., 2021; M. Zhu et al., 2022; M. Yang et al., 2023; J. Yan et al., 2022; F. Song et al., 2023; G. Pasini et al., 2023; H. Liu et al., 2021; R. Han et al., 2020; Y.X. Guo et al., 2021; Z.Y. Chen et al., 2022) used the internal validation method of random sampling, 12 articles (H.H. Li et al., 2021; C. Alvarez-Jimenez et al., 2020; B. Dunn et al., 2023; H. Liu et al., 2019; R. Patil et al., 2016; D.D. Yu et al., 2017; F.C. Yang et al., 2021; J. Liu et al., 2019; E. Linning et al., 2019; Z. Khodabakhshi et al., 2021; Q.B. Gu et al., 2019; L.N. E et al., 2019) used the k-fold cross-validation method, and 4 articles (A. Brunetti et al., 2022; F. Song et al., 2023; Y. Dong et al., 2022; X.J. Chu et al., 2023) used the external validation method. Among the models, the most common ones were logistic regression, support vector machine, and random forest models (Table S2).

4.3 Assessment of Study Quality

Article quality was evaluated using the RQS scale. Seven studies did not mention imaging protocols and therefore did not score in this regard. None of the studies conducted pre-tests with different parameters and repeated measures tests with different parameters, so none of them scored. In terms of the number of participants of the imager, 19 articles explicitly described multiple participants, and 10 articles did not explicitly describe the number of participants involved in the image segmentation, and four articles described that only one person performed the segmentation, so these articles were not scored in this regard. Regarding the validation set, 6 articles were externally validated in multiple centers, so they were scored 4-5. Furthermore, some included models lacked a region of interest (ROI) region segmentation protocol for images, as well as code for texture extraction and modeling. Consequently, no score was given for these entries. Finally, the mean score for enrolled articles was 7.12.

4.4 Meta-Analysis

4.4.1 Ki-67

1) Synthesized Results

A random-effects model was employed for meta-analysis to identify Ki-67 expression. In the training set, the ML model for Ki-67 levels in lung cancer showed 0.82 c-index (95% CI: 0.78-0.85). The clinical features-based ML models manifested 0.75 c-index (95%CI: 0.67-0.82), the model based on radiomics features demonstrated 0.84 c-index (95%CI: 0.80-0.88), and the model of radiomics combined with clinical features yielded 0.83 c-index (95%CI: 0.74-0.93) (Figure 2).

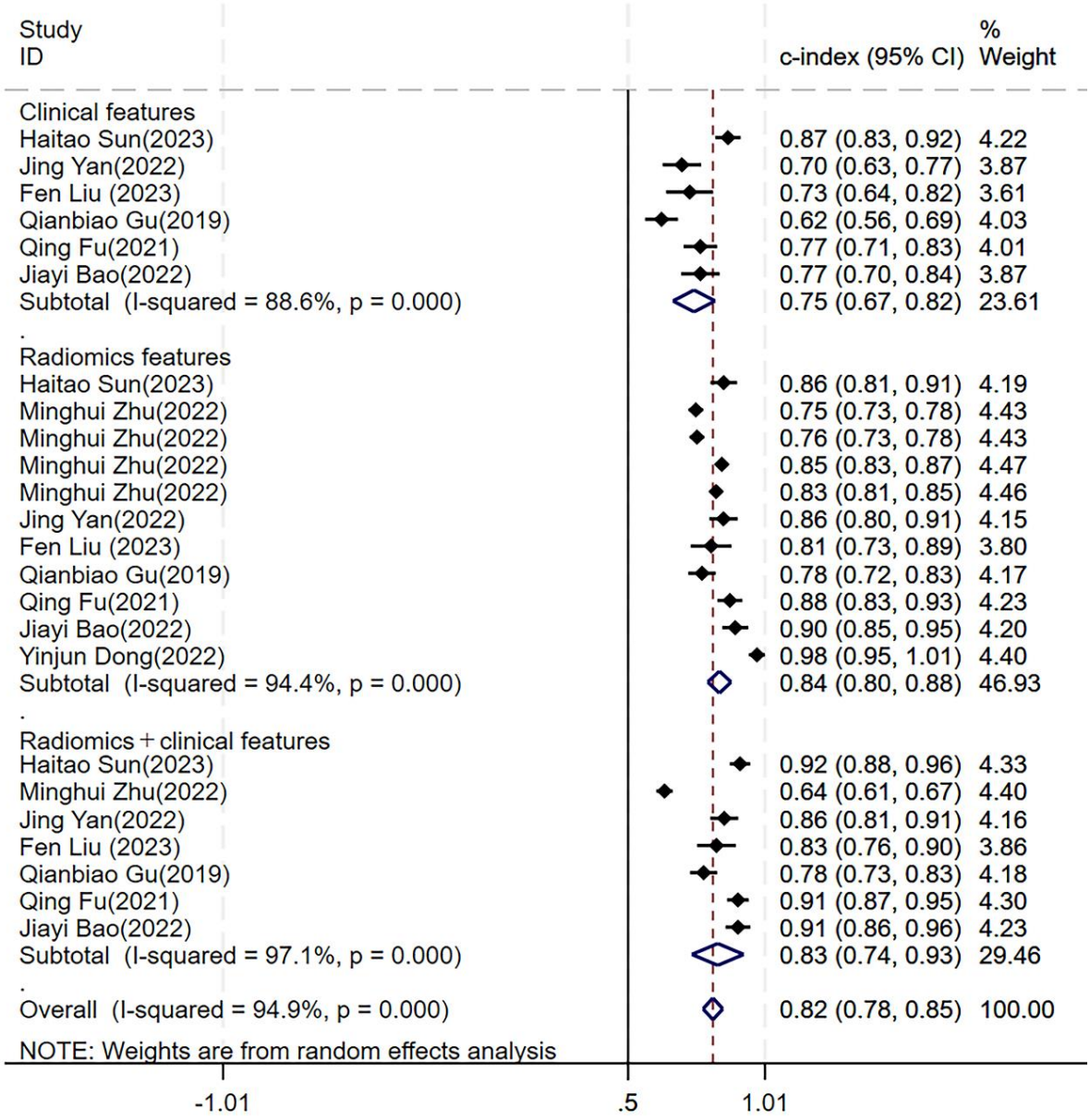


Figure 2. Forest plot of radiomics prediction of LADC Ki-67 meta-analysis in the test sets

In the training set, the model based on clinical features alone yielded 0.70 sensitivity (95% CI: 0.64-0.76) and 0.74 specificity (95% CI: 0.57-0.86). Radiomics-based models demonstrated 0.80 sensitivity (95% CI: 0.76-0.84) and 0.75 specificity (95% CI: 0.69-0.80). The model incorporated radiomics and clinical features demonstrated 0.77 sensitivity (95% CI: 0.71-0.81) and 0.82 specificity (95% CI: 0.73-0.88) (Figure S1-3).

In the validation set, the ML model in forecasting Ki-67 levels in lung cancer exhibited 0.77 c-index (95% CI: 0.74-0.79). The clinical features-based ML models showcased 0.72 c-index (95% CI: 0.68-0.75) radiomics-based model displayed 0.76 c-index (95% CI: 0.72-0.80), while radiomics + clinical features model manifested 0.81 c-index (95% CI: 0.75-0.87) (Figure 3).

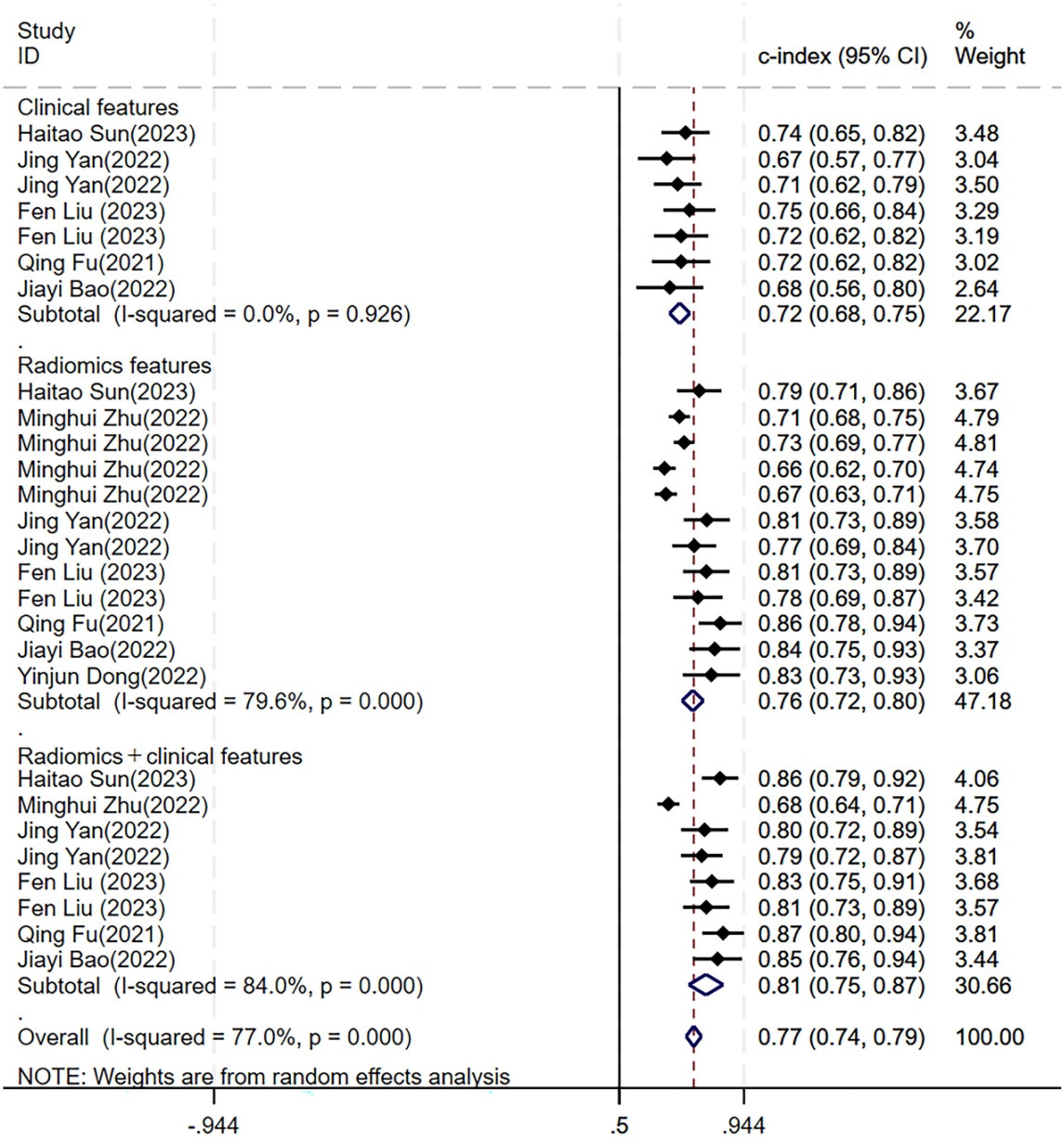


Figure 3. Forest plot of radiomics prediction of LADC ki-67 meta-analysis in the validation sets

In the validation set, the clinical features-based model yielded 0.68 sensitivity (95% CI: 0.58-0.76) and 0.70 specificity (95% CI: 0.63-0.76). The radiomics-based model exhibited 0.76 sensitivity (95% CI: 0.71-0.81) and 0.74 specificity (95% CI: 0.67-0.80). The combined model demonstrated 0.79 sensitivity (95% CI: 0.72-0.84) and 0.78 specificity (95% CI: 0.71-0.84) (Figure S4-6).

2) Reporting Biases

In the training set, the funnel plots showed no publication bias, and Egger's test revealed P= 0.849 (Figure S7).

In the validation set, funnel plots showed a publication bias, and Egger's test revealed P= 0.001 (Figure S8).

4.5 Histological Subtypes

4.5.1 Synthesized Results

In all included articles, radiomics was employed solely to identify the histological subtypes. In the training set, the radiomics-based pooling demonstrated 0.87 c-index (95% CI: 0.86-0.89) (Figure 4), with 0.82 sensitivity (95% CI: 0.75-0.87) and 0.69 specificity (0.61-0.75) (Figure S9). Additionally, logistic regression and support vector machine showed c-indexes of 0.89 (95% CI: 0.85-0.92) and 0.82 (95% CI: 0.85-0.92), respectively. In the validation set, the c-index based on radiomics pooling was 0.78 (95% CI: 0.76-0.80) (Figure 5), with 0.78 sensitivity (95% CI: 0.70-0.84) and 0.79 specificity (95% CI: 0.73-0.85) (Figure S10).

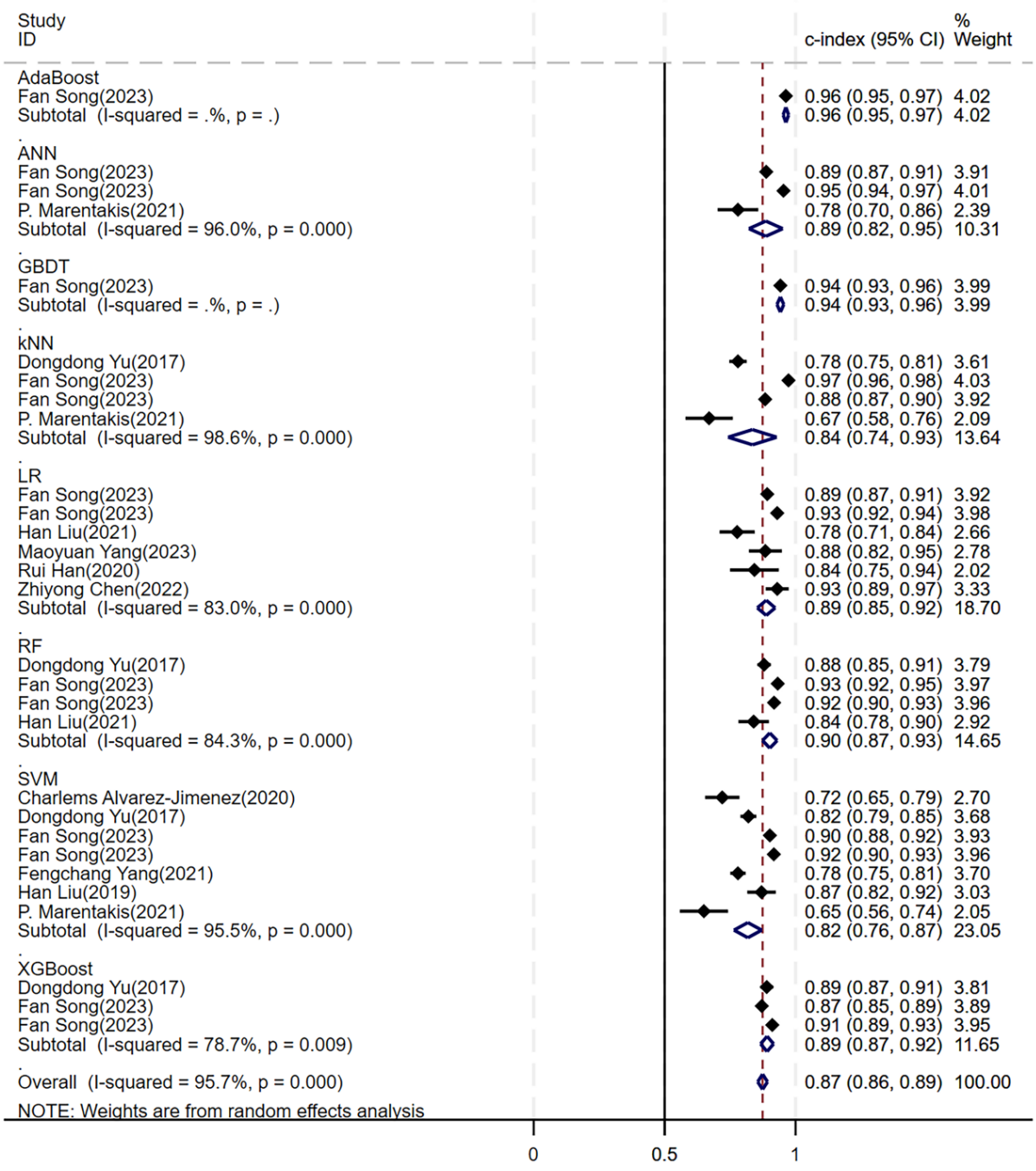


Figure 4. Forest plot of histological subtypes of LADC predicted by radiomics in the test sets

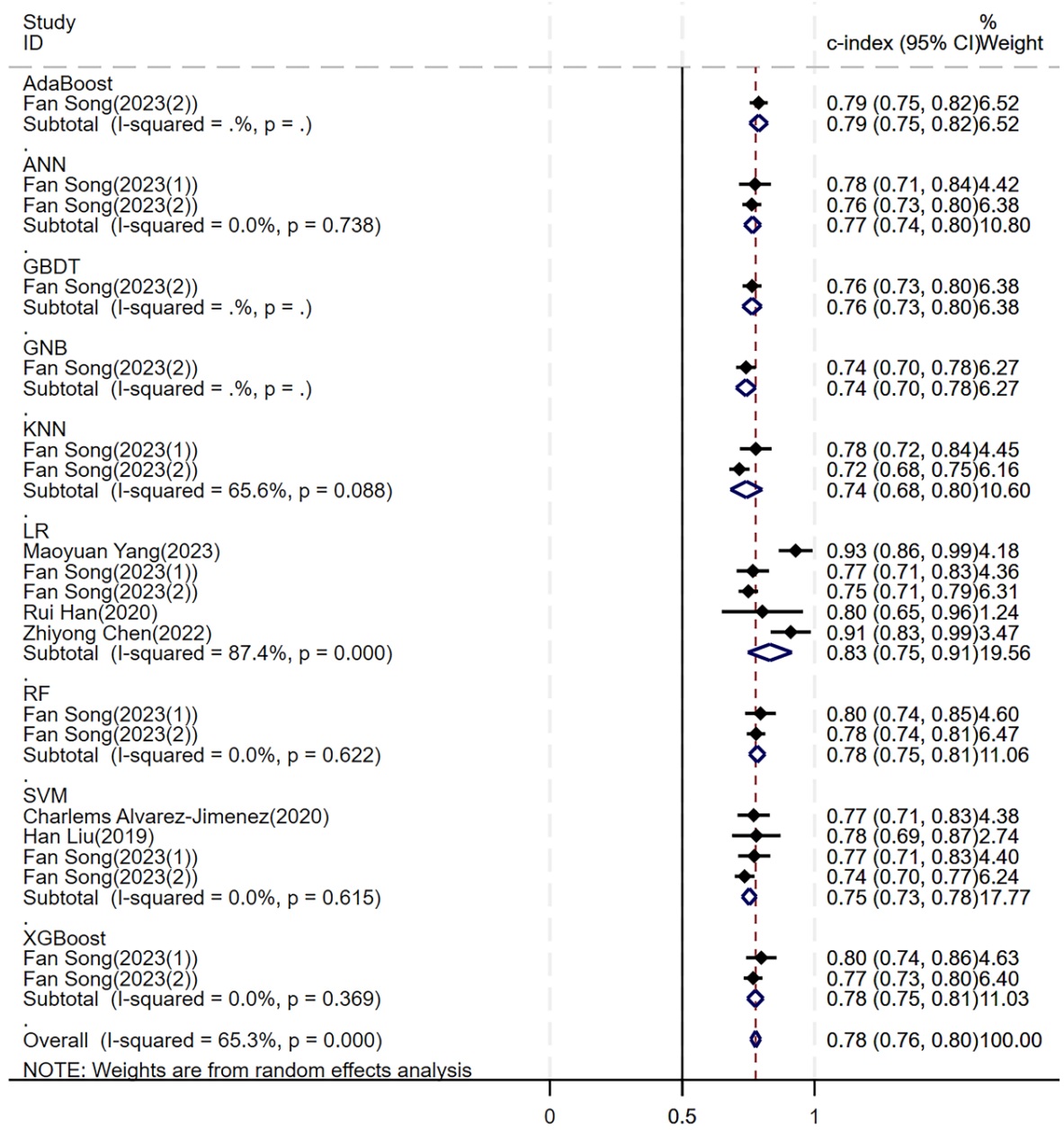


Figure 5. Forest plot of histological subtypes of LADC predicted by radiomics in the validation sets

4.5.2 Reporting Biases

In the training set, the funnel plots showed a publication bias, and Egger's test revealed P = 0.0001 (Figure S11).

In the validation set, funnel plots noted a publication bias, and Egger's test revealed P = 0.034 (Figure S12).

4.6 Multi-classification of Pathological Types

In our study, 12 articles employed a multiclassification approach to identify distinct pathological subtypes, with 3 presenting confusion matrices. The types identified in these 3 articles were predominantly LADC, LSCC, LCC, LASC, and NOS. The discriminatory accuracy of LADC was 0.76, 0.60, 0.58, 0.71, SCC was 0.94, 0.81, 0.78, 0.76, LCC was 0.78, 0.73, 0.75, and LASC was 0.67, respectively. However, the other articles did not provide multiclassification accuracies, so they cannot be discussed further here.

5. Discussion

5.1 Summary of the Main Findings

ML is effective in identifying Ki-67 levels in lung cancer and in differentiating between subtypes. In the validation set, the c-index for Ki-67 expression was 0.76 (95% CI: 0.72-0.80) for radiomics and 0.81 (95% CI:

0.75-0.87) for radiomics + clinical features. In histological subtypes, the included studies constructed ML models for identifying histological subtypes based on the radiomics approach, but there were no studies on clinical features or clinical features + radiomics to construct models. The c-index for the identification of pathotypes was 0.78 (95% CI: 0.76-0.80).

5.2 Comparison with Other Reviews

Research has investigated the potential of a preoperative test for Ki-67 and histological subtypes. Wei et al. (2018) unveiled that Ki-67 elevation was correlated with gender, age, smoking, tumor size, and pathological stage. Liu et al. (2023) demonstrated that nomograms integrating imaging and histological features with clinical characteristics may represent a promising non-invasive approach for forecasting Ki-67 levels in individuals with pure solid NSCLC. Zhu et al. (2022) initially investigated the power of intra- and perinodal radiographic features in forecasting Ki-67 levels, with an AUC of 0.731 (0.662-0.799). While Yao et al. (2022) indicated that 18F-FDG PET/CT-based imaging histological features demonstrated excellent performance in predicting Ki67 expression, with an AUC of 0.85 (95% CI, 0.71-0.98), 0.83 accuracy (95% CI, 0.66-0.93), 0.94 sensitivity, and 0.72 specificity in the test set.

In terms of histological subtypes, most studies (H. Liu et al., 2019; X. Tang et al., 2022; D.D. Yu et al., 2017; M. Saad & T.S. Choi, 2018; X. Zhu et al., 2018) concentrated on the classification of LSCC and LADC and indicated that radiomics may be a valuable approach for differentiating between these two types before biopsy and surgery. Saad et al. (2018) proposed a computational method that combined computerized subtyping and prognosis, which was effective in subtyping patients with LADC and LSCC with an accuracy of 73.3% to 93%. Zhu et al. (2018) developed an imaging histological signature comprising 5 quantitative CT image features to differentiate ADC from SCC preoperatively. This approach exhibited satisfactory performance in both the validation and training sets, with 0.828 sensitivity and 0.900 specificity in the validation set.

Furthermore, lung cancer encompasses diverse histological subtypes, including LCC, LASC, and NOS. These different types exhibit varying treatment and prognosis profiles, underscoring the challenges in distinguishing between LADC and LSCC distinction. Consequently, some researchers have conducted studies on classifying lung cancer into multiple categories. Chu et al. (2023) used an optimal three-class classification model to distinguish LASC from LADC or LSCC and achieved an AUC of 0.89 and an accuracy of 0.81 in an external validation set. Dunn et al. (2023) utilized clinical CT/PET images for the first to classify 3 histological subtypes of lung cancer, with an accuracy of 92.7% and an AUC of 0.97. Consequently, further studies on multiple classifications of lung cancer are required to provide greater clinical value.

Clinical features are pivotal in the identification of Ki-67. Our study sought to ascertain the impact of clinical features, radiomics features, and radiomics + clinical features in identifying Ki-67. The findings indicate that radiomics + clinical features may provide increased predictive precision in identifying Ki-67 expression. Nevertheless, these results are based on limited evidence, highlighting the necessity for additional research in future investigations. Moreover, clinical characteristics were not identified for histological subtypes, possibly due to the lack of clear distinctions between different stages of disease progression.

5.3 Advantages and Limitations

This paper presents the initial investigation into the value of radiomics for Ki-67 expression and histological subtypes. However, this study has the following limitations: Firstly, about Ki-67 expression, there may be some differences in the cut-off thresholds across studies. Due to the limited included studies, we failed to discuss the impact of these differences on model construction. Secondly, in the identification of Ki-67, we did not further discuss the predictive power of the models with different variables owing to the limited number of models. Thirdly, the included articles only employed ML models constructed by radiomics for pathotype identification. Consequently, the original studies failed to acknowledge the crucial role of modeling variables in pathotyping, resulting in a lack of further discussion in our review. Fourthly, the models included were primarily validated through random sampling methods, with a notable absence of external validation from multiple centers.

6. Conclusions

Radiomics serves as a valuable instrument for the non-invasive detection of high Ki-67 levels in lung cancer, showcasing significant accuracy in identifying different lung cancer pathotypes. However, the current findings are grounded on restricted evidence, presenting a notable risk of bias in certain studies that have employed radiomics. Therefore, future research endeavors should prioritize the validation of these findings through larger samples and the development of an intelligent readable tool.

Declaration of Interests

The authors declare that they have no known competing financial interests or personal relationships that could

have appeared to influence the work reported in this paper.

Informed Consent and Patient Details

Not applicable.

Abbreviations

CI, confidence intervals; LADC, lung adenocarcinoma; LSCC, squamous cell carcinoma; LASC, adeno-squamous carcinoma; ML, machine learning; NSCLC, non-small cell lung cancer; NOS, not otherwise specified; SCLC, small cell lung cancer; ROI, region of interest.

References

A. Brunetti, N. Altini, D. Buongiorno, E. Garolla, F. Corallo, M. Gravina, V. Bevilacqua and B. Prencipe, (2022). A Machine Learning and Radiomics Approach in Lung Cancer for Predicting Histological Subtype. *Applied Sciences-Basel*, 12. https://doi.org/10.3390/app12125829.

A. Haga, W. Takahashi, S. Aoki, K. Nawa, H. Yamashita, O. Abe and K. Nakagawa, (2018). Classification of early-stage non-small cell lung cancers on computed tomographic images into histological types using radiomic features: interobserver delineation variability analysis. *Radiological Physics and Technology*, *11*, 27-35. https://doi.org/10.1007/s12194-017-0433-2.

A. Warth, J. Cortis, A. Soltermann, M. Meister, J. Budczies, A. Stenzinger, B. Goeppert, M. Thomas, F.J. Herth, P. Schirmacher, P.A. Schnabel, H. Hoffmann, H. Dienemann, T. Muley and W. Weichert, (2014). Tumour cell proliferation (Ki-67) in non-small cell lung cancer: a critical reappraisal of its prognostic role. *Br J Cancer*, *111*, 1222-9. https://doi.org/10.1038/bjc.2014.402.

B. Dunn, M. Pierobon and Q. Wei, (2023). Automated Classification of Lung Cancer Subtypes Using Deep Learning and CT-Scan Based Radiomic Analysis. *Bioengineering (Basel)*, 10. https://doi.org/10.3390/bioengineering10060690.

C. Alvarez-Jimenez, A.A. Sandino, P. Prasanna, A. Gupta, S.E. Viswanath and E. Romero, (2020). Identifying cross-scale associations between radiomic and pathomic signatures of non-small cell lung cancer subtypes: Preliminary results. *Cancers*, *12*, 1-17. https://doi.org/10.3390/cancers12123663.

D.D. Yu, Y.L. Zang, D. Dong, M. Zhou, O. Gevaert, M.J. Fang, J.Y. Shi and J. Tian, (2017). Developing a Radiomics Framework for Classifying Non-Small Cell Lung Carcinoma Subtypes, Conference on Medical Imaging — Computer-Aided Diagnosis, Orlando, FL.

D.M. Wei, W.J. Chen, R.M. Meng, N. Zhao, X.Y. Zhang, D.Y. Liao and G. Chen, (2018). Augmented expression of Ki-67 is correlated with clinicopathological characteristics and prognosis for lung cancer patients: an up-dated systematic review and meta-analysis with 108 studies and 14,732 patients. *Respir Res*, *19*, 150. https://doi.org/10.1186/s12931-018-0843-7.

E. Linning, L. Lu, L. Li, H. Yang, L.H. Schwartz and B. Zhao, (2019). Radiomics for Classifying Histological Subtypes of Lung Cancer Based on Multiphasic Contrast-Enhanced Computed Tomography. J*ournal of Computer Assisted Tomography*, *43*, 300-306. https://doi.org/10.1097/RCT.0000000000000836.

F. Bray, M. Laversanne, H. Sung, J. Ferlay, R.L. Siegel, I. Soerjomataram and A. Jemal, (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, *74*(3), 229-263. https://doi.org/10.3322/caac.21834.

F. Liu, Q. Li, Z. Xiang, X. Li, F. Li, Y. Huang, Y. Zeng, H. Lin, X. Fang and Q. Yang, (2023). CT radiomics model for predicting the Ki-67 proliferation index of pure-solid non-small cell lung cancer: a multicenter study. *Front Oncol*, *13*, 1175010. https://doi.org/10.3389/fonc.2023.1175010.

F. Liu, Q.C. Li, Z.Q. Xiang, X.F. Li, F.T. Li, Y.Q. Huang, Y. Zeng, H.S. Lin, X.J. Fang and Q.L. Yang, (2023). CT radiomics model for predicting the Ki-67 proliferation index of pure-solid non-small cell lung cancer: a multicenter study. *Frontiers in Oncology*, 13. https://doi.org/10.3389/fonc.2023.1175010.

F. Song, J. Tian, P. Zhang, C. Ma, Y. Sun, Y. Feng, T. Zhang, Y. Lei, Y. He, Z. Cai, Y. Cheng and G. Zhang, (2023). A Novel Feature Engineering Method Based on Latent Representation Learning for Radiomics: Application in NSCLC Subtype Classification. *IEEE Journal of Biomedical and Health Informatics*, 1-11. https://doi.org/10.1109/JBHI.2023.3290006.

F. Song, X. Song, Y.D. Feng, G.D. Fan, Y.Y. Sun, P. Zhang, J.K. Li, F. Liu and G.L. Zhang, (2023). Radiomics feature analysis and model research for predicting histopathological subtypes of non-small cell lung cancer on CT images: A multi-dataset study. *Medical Physics*, *50*, 4351-4365. https://doi.org/10.1002/mp.16233.

F.C. Yang, W. Chen, H.F. Wei, X.R. Zhang, S.H. Yuan, X. Qiao and Y.W. Chen, (2021). Machine Learning for Histologic Subtype Classification of Non-Small Cell Lung Cancer: A Retrospective Multicenter Radiomics

Study. *Frontiers in Oncology*, 10. https://doi.org/10.3389/fonc.2020.608598.

G. Pasini, A. Stefano, G. Russo, A. Comelli, F. Marinozzi and F. Bini, (2023). Phenotyping the Histopathological Subtypes of Non-Small-Cell Lung Carcinoma: How Beneficial Is Radiomics? *Diagnostics*, 13. https://doi.org/10.3390/diagnostics13061167.

H. Liu, B. Jing, W.J. Han, Z.Q. Long, X. Mo and H.Y. Li, (2019). A Comparative Texture Analysis Based on NECT and CECT Images to Differentiate Lung Adenocarcinoma from Squamous Cell Carcinoma. *Journal of Medical Systems*, 43. https://doi.org/10.1007/s10916-019-1175-y.

H. Liu, Z.C. Jiao, W.J. Han and B. Jing, (2021). Identifying the histologic subtypes of non-small cell lung cancer with computed tomography imaging: a comparative study of capsule net, convolutional neural network, and radiomics. *Quantitative Imaging in Medicine and Surgery*, *11*, 2756-2765. https://doi.org/10.21037/qims-20-734.

H. Sun, P. Zhou, G. Chen, Z. Dai, P. Song and J. Yao, (2023). Radiomics nomogram for the prediction of Ki-67 index in advanced non-small cell lung cancer based on dual-phase enhanced computed tomography. *Journal of Cancer Research and Clinical Oncology*, *149*, 9301-9315. https://doi.org/10.1007/s00432-023-04856-2.

H.H. Li, L. Gao, H. Ma, D. Arefan, J.C.A. He, J.Q. Wang and H. Liu, Radiomics-Based Features for Prediction of Histological Subtypes in Central Lung Cancer. *Frontiers in Oncology*, 11. https://doi.org/10.3389/fonc.2021.658887.

J. Bao, Y. Liu, X. Ping, X. Zha, S. Hu and C. Hu, (2022). Preoperative Ki-67 proliferation index prediction with a radiomics nomogram in stage T1a-b lung adenocarcinoma. *European Journal of Radiology*, 155. https://doi.org/10.1016/j.ejrad.2022.110437.

J. Lin, Y.J. Yu, X.L. Zhang, Z.L. Wang and S.J. Li, (2023). Classification of Histological Types and Stages in Non-small Cell Lung Cancer Using Radiomic Features Based on CT Images. *Journal of Digital Imaging*, *36*, 1029-1037. https://doi.org/10.1007/s10278-023-00792-2.

J. Liu, J.J. Cui, F. Liu, Y.X. Yuan, F. Guo and G.L. Zhang, (2019). Multi-subtype classification model for non-small cell lung cancer based on radiomics: SLS model. *Medical Physics*, *46*, 3091-3100. https://doi.org/10.1002/mp.13551.

J. Yan, X. Xue, C. Gao, Y.F. Guo, L.Y. Wu, C.Y. Zhou, F. Chen and M.S. Xu, (2022). Predicting the Ki-67 proliferation index in pulmonary adenocarcinoma patients presenting with subsolid nodules: construction of a nomogram based on CT images. *Quantitative Imaging in Medicine and Surgery*, *12*, 642-+. https://doi.org/10.21037/qims-20-1385.

J.S. Donington, Y.T. Kim, B. Tong, A.L. Moreira, J. Bessich, K.D. Weiss, Y.L. Colson, D. Wigle, R.U. Osarogiagbon, J. Zweig, H. Wakelee, J. Blasberg, M. Daly, L. Backhus and P. Van Schil, (2018). Progress in the Management of Early-Stage Non-Small Cell Lung Cancer in 2017. *J Thorac Oncol*, *13*, 767-778. https://doi.org/10.1016/j.jtho.2018.04.002.

L. Crinò, W. Weder, J. van Meerbeeck and E. Felip, (2010). Early stage and locally advanced (non-metastatic) non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, *21*(Suppl 5), v103-15. https://doi.org/10.1093/annonc/mdq207.

L.N. E, L. Lu, L. Li, H. Yang, L.H. Schwartz and B.S. Zhao, (2019). Radiomics for Classification of Lung Cancer Histological Subtypes Based on Nonenhanced Computed Tomography. *Academic Radiology*, *26*, 1245-1252. https://doi.org/10.1016/j.acra.2018.10.013.

M. Jamal-Hanjani, S.A. Quezada, J. Larkin and C. Swanton, (2015). Translational implications of tumor heterogeneity. *Clin Cancer Res*, *21*, 1258-66. https://doi.org/10.1158/1078-0432.Ccr-14-1429.

M. Saad, T.S. Choi, (2018). Computer-assisted subtyping and prognosis for non-small cell lung cancer patients with unresectable tumor. *Comput Med Imaging Graph*, *67*, 1-8. https://doi.org/10.1016/j.compmedimag.2018.04.003.

M. Yang, L. Shi, T. Huang, G. Li, H. Shao, Y. Shen, J. Zhu and B. Ni, (2023). Value of contrast-enhanced magnetic resonance imaging-T2WI-based radiomic features in distinguishing lung adenocarcinoma from lung squamous cell carcinoma with solid components >8 mm. *Journal of Thoracic Disease*, *15*, 635-648. https://doi.org/10.21037/jtd-23-142.

M. Zhu, Z. Yang, W. Zhao, M. Wang, W. Shi, Z. Cheng, C. Ye, Q. Zhu, L. Liu, Z. Liang and L. Chen, (2022). Predicting Ki-67 labeling index level in early-stage lung adenocarcinomas manifesting as ground-glass opacity nodules using intra-nodular and peri-nodular radiomic features. *Cancer Medicine*, *11*, 3982-3992. https://doi.org/10.1002/cam4.4719.

M.K. Kamel, B. Lee, S.W. Harrison, J.L. Port, N.K. Altorki and B.M. Stiles, (2022). Sublobar resection is comparable to lobectomy for screen-detected lung cancer. *J Thorac Cardiovasc Surg*, *163*, 1907-1915. https://doi.org/10.1016/j.jtcvs.2021.06.056.

M.S. Tsao, A.G. Nicholson, J.J. Maleszewski, A. Marx and W.D. Travis, (2022). Introduction to 2021 WHO Classification of Thoracic Tumors. *J Thorac Oncol*, *17*, e1-e4. https://doi.org/10.1016/j.jtho.2021.09.017.

N. Duma, R. Santana-Davila and J.R. Molina, (2019). Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clin Proc*, *94*, 1623-1640. https://doi.org/10.1016/j.mayocp.2019.01.013.

P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G. van Stiphout, P. Granton, C.M. Zegers, R. Gillies, R. Boellard, A. Dekker and H.J. Aerts, (2012). Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*, *48*, 441-6. https://doi.org/10.1016/j.ejca.2011.11.036.

P. Marentakis, P. Karaiskos, V. Kouloulias, N. Kelekis, S. Argentos, N. Oikonomopoulos and C. Loukas, (2021). Lung cancer histology classification from CT images based on radiomics and deep learning models. *Medical and Biological Engineering and Computing*, *59*, 215-226. https://doi.org/10.1007/s11517-020-02302-w.

Q. Fu, S.L. Liu, D.P. Hao, Y.B. Hu, X.J. Liu, Z. Zhang, W.H. Wang, X.Y. Tang, C.Y. Zhang and S.H. Liu, (2021). CT Radiomics Model for Predicting the Ki-67 Index of Lung Cancer: An Exploratory Study. *Front Oncol*, *11*, 743490. https://doi.org/10.3389/fonc.2021.743490.

Q.B. Gu, Z.C. Feng, Q. Liang, M.J. Li, J. Deng, M.T. Ma, W. Wang, J.B. Liu, P. Liu and P.F. Rong, (2019). Machine learning-based radiomics strategy for prediction of cell proliferation in non-small cell lung cancer. *European Journal of Radiology*, *118*, 32-37. https://doi.org/10.1016/j.ejrad.2019.06.025.

R. Han, R. Arjal, J. Dong, H. Jiang, H. Liu, D.Y. Zhang and L. Huang, (2020). Three dimensional texture analysis of noncontrast chest CT in differentiating solitary solid lung squamous cell carcinoma from adenocarcinoma and correlation to immunohistochemical markers. *Thoracic Cancer*, *11*, 3099-3106. https://doi.org/10.1111/1759-7714.13592.

R. Patil, G. Mahadevaiah and A. Dekker, (2016). An Approach Toward Automatic Classification of Tumor Histopathology of Non-Small Cell Lung Cancer Based on Radiomic Features. *Tomography*, *2*, 374-377. https://doi.org/10.18383/j.tom.2016.00244.

T.P. Debray, J.A. Damen, R.D. Riley, K. Snell, J.B. Reitsma, L. Hooft, G.S. Collins and K.G. Moons, (2019). 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.

W. Yao, Y. Liao, X. Li, F. Zhang, H. Zhang, B. Hu, X. Wang, L. Li and M. Xiao, (2022). Noninvasive Method for Predicting the Expression of Ki67 and Prognosis in Non-Small-Cell Lung Cancer Patients: Radiomics. *J Healthc Eng*, 2022, 7761589. https://doi.org/10.1155/2022/7761589.

W.D. Travis, H. Asamura, A.A. Bankier, M.B. Beasley, F. Detterbeck, D.B. Flieder, J.M. Goo, H. MacMahon, D. Naidich, A.G. Nicholson, C.A. Powell, M. Prokop, R. Rami-Porta, V. Rusch, P. van Schil and Y. Yatabe, (2016). The IASLC Lung Cancer Staging Project: Proposals for Coding T Categories for Subsolid Nodules and Assessment of Tumor Size in Part-Solid Tumors in the Forthcoming Eighth Edition of the TNM Classification of Lung Cancer. *J Thorac Oncol*, *11*, 1204-1223. https://doi.org/10.1016/j.jtho.2016.03.025.

X. Chu, L. Niu, X. Yang, S. He, A. Li, L. Chen, Z. Liang, D. Jing and R. Zhou, (2023). Radiomics and deep learning models to differentiate lung adenosquamous carcinoma: A multicenter trial. *iScience*, *26*, 107634. https://doi.org/10.1016/j.isci.2023.107634.

X. Tang, H.L. Huang, P. Du, L.J. Wang, H. Yin and X.P. Xu, (2022). Intratumoral and peritumoral CT-based radiomics strategy reveals distinct subtypes of non-small-cell lung cancer. *Journal of Cancer Research and Clinical Oncology*, *148*, 2247-2260. https://doi.org/10.1007/s00432-022-04015-z.

X. Zhu, D. Dong, Z. Chen, M. Fang, L. Zhang, J. Song, D. Yu, Y. Zang, Z. Liu, J. Shi and J. Tian, (2018). Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *Eur Radiol*, *28*, 2772-2778. https://doi.org/10.1007/s00330-017-5221-1.

X.J. Chu, L.S. Niu, X.H. Yang, S.Q. He, A.X. Li, L. Chen, Z. Liang, D. Jing and R.R. Zhou, (2023). Radiomics and deep learning models to differentiate lung adenosquamous carcinoma: A multicenter trial. *Iscience*, 26. https://doi.org/10.1016/j.isci.2023.107634.

Y. Dong, Z. Jiang, C. Li, S. Dong, S. Zhang, Y. Lv, F. Sun and S. Liu, (2022). Development and validation of novel radiomics-based nomograms for the prediction of EGFR mutations and Ki-67 proliferation index in non-small cell lung cancer. *Quantitative Imaging in Medicine and Surgery*, *12*, 2658-2671.

https://doi.org/10.21037/qims-21-980.

Y.X. Guo, Q. Song, M.M. Jiang, Y.L. Guo, P. Xu, Y.Q. Zhang, C.C. Fu, Q. Fang, M.S. Zeng and X.Z. Yao, (2021). Histological Subtypes Classification of Lung Cancers on CT Images Using 3D Deep Learning and Radiomics. *Academic Radiology*, *28*, E258-E266. https://doi.org/10.1016/j.acra.2020.06.010.

Z. Khodabakhshi, S. Mostafaei, H. Arabi, M. Oveisi, I. Shiri and H. Zaidi, (2021). Non-small cell lung carcinoma histopathological subtype phenotyping using high-dimensional multinomial multiclass CT radiomics signature. *Computers in Biology and Medicine*, 136, https://doi.org/10.1016/j.compbiomed.2021.104752.

Z. Li, F. Li, C. Pan, Z. He, X. Pan, Q. Zhu, W. Wu and L. Chen, (2021). Tumor cell proliferation (Ki-67) expression and its prognostic significance in histological subtypes of lung adenocarcinoma. *Lung Cancer*, *154*, 69-75. https://doi.org/10.1016/j.lungcan.2021.02.009.

Z.Y. Chen, L. Yi, Z.W. Peng, J.Z. Zhou, Z.T. Zhang, Y.H. Tao, Z. Lin, A.J. He, M.N. Jin and M.J. Zuo, (2022). Development and validation of a radiomic nomogram based on pretherapy dual-energy CT for distinguishing adenocarcinoma from squamous cell carcinoma of the lung. *Frontiers in Oncology*, 12. https://doi.org/10.3389/fonc.2022.949111.

Appendix

Table S1. Literature search strategy

1) Pubmed

Search number	Query	Results
#1	“Lung Neoplasms” [Mesh]	278,964
#2	((((((((((((((((((((((((((((Pulmonary Neoplasms[Title/Abstract]) OR (Lung Neoplasm[Title/Abstract])) OR (Pulmonary Neoplasm[Title/Abstract])) OR (Lung Cancer[Title/Abstract])) OR (Lung Cancers[Title/Abstract])) OR (Pulmonary Cancer[Title/Abstract])) OR (Lung Carcinoma[Title/Abstract])) OR (Lung Carcinomas[Title/Abstract])) OR (Pulmonary Cancer[Title/Abstract])) OR (Pulmonary Cancers[Title/Abstract])) OR (Cancer of the Lung[Title/Abstract])) OR (Cancer of Lung[Title/Abstract])) OR (Adenocarcinoma of Lung[Title/Abstract])) OR (Lung Adenocarcinomas[Title/Abstract])) OR (Lung Adenocarcinoma[Title/Abstract])) OR (Small Cell Cancer Of The Lung[Title/Abstract])) OR (Oat Cell Carcinoma of Lung[Title/Abstract])) OR (lung tumor[Title/Abstract])) OR (pulmonary tumour[Title/Abstract])) OR (pulmonary tumor[Title/Abstract])) OR (broncho-pulmonary neoplasm[Title/Abstract])) OR (broncho-pulmonary tumor[Title/Abstract])) OR (bronchopulmonary neoplasia[Title/Abstract])) OR (bronchopulmonary neoplasm[Title/Abstract])) OR (bronchopulmonary tumor[Title/Abstract])) OR (lung neoplasia[Title/Abstract])) OR (lung tumour[Title/Abstract])) OR (pulmonary neoplasia[Title/Abstract])) OR (broncho-pulmonary cancer[Title/Abstract])) OR (bronchopulmonary cancer[Title/Abstract])) OR (NSCLC[Title/Abstract])	251,435
#3	(“Lung Neoplasms”[Mesh]) OR ((((((((((((((((((((((((((((Pulmonary Neoplasms[Title/Abstract]) OR (Lung Neoplasm[Title/Abstract])) OR (Pulmonary Neoplasm[Title/Abstract])) OR (Lung Cancer[Title/Abstract])) OR (Lung Cancers[Title/Abstract])) OR (Pulmonary Cancer[Title/Abstract])) OR (Lung Carcinoma[Title/Abstract])) OR (Lung Carcinomas[Title/Abstract])) OR (Pulmonary Cancer[Title/Abstract])) OR (Pulmonary Cancers[Title/Abstract])) OR (Cancer of the Lung[Title/Abstract])) OR (Cancer of Lung[Title/Abstract])) OR (Adenocarcinoma of Lung[Title/Abstract])) OR (Lung Adenocarcinomas[Title/Abstract])) OR (Lung Adenocarcinoma[Title/Abstract])) OR (Small Cell Cancer Of The Lung[Title/Abstract])) OR (Oat Cell Carcinoma of Lung[Title/Abstract])) OR (lung tumor[Title/Abstract])) OR (pulmonary tumour[Title/Abstract])) OR (pulmonary tumor[Title/Abstract])) OR (broncho-pulmonary neoplasm[Title/Abstract])) OR (broncho-pulmonary tumor[Title/Abstract])) OR (bronchopulmonary neoplasia[Title/Abstract])) OR (bronchopulmonary neoplasm[Title/Abstract])) OR (bronchopulmonary	363,334

	tumor[Title/Abstract])) OR (lung neoplasia[Title/Abstract])) OR (lung tumour[Title/Abstract])) OR (pulmonary neoplasia[Title/Abstract])) OR (broncho-pulmonary cancer[Title/Abstract])) OR (bronchopulmonary cancer[Title/Abstract])) OR (NSCLC[Title/Abstract]))	
#4	((((Radiomics[Title/Abstract]) OR (radiomic[Title/Abstract])) OR (radiogenomic[Title/Abstract])) OR (radiomics-based[Title/Abstract])) OR (Texture[Title/Abstract])	50,443
#5	(("Lung Neoplasms"[Mesh]) OR ((((((((((((((((((((((((((Pulmonary Neoplasms[Title/Abstract]) OR (Lung Neoplasm[Title/Abstract])) OR (Pulmonary Neoplasm[Title/Abstract])) OR (Lung Cancer[Title/Abstract])) OR (Lung Cancers[Title/Abstract])) OR (Pulmonary Cancer[Title/Abstract])) OR (Lung Carcinoma[Title/Abstract])) OR (Lung Carcinomas[Title/Abstract])) OR (Pulmonary Cancer[Title/Abstract])) OR (Pulmonary Cancers[Title/Abstract])) OR (Cancer of the Lung[Title/Abstract])) OR (Cancer of Lung[Title/Abstract])) OR (Adenocarcinoma of Lung[Title/Abstract])) OR (Lung Adenocarcinomas[Title/Abstract])) OR (Lung Adenocarcinoma[Title/Abstract])) OR (Small Cell Cancer Of The Lung[Title/Abstract])) OR (Oat Cell Carcinoma of Lung[Title/Abstract])) OR (lung tumor[Title/Abstract])) OR (pulmonary tumour[Title/Abstract])) OR (pulmonary tumor[Title/Abstract])) OR (broncho-pulmonary neoplasm[Title/Abstract])) OR (broncho-pulmonary tumor[Title/Abstract])) OR (bronchopulmonary neoplasia[Title/Abstract])) OR (bronchopulmonary neoplasm[Title/Abstract])) OR (bronchopulmonary tumor[Title/Abstract])) OR (lung neoplasia[Title/Abstract])) OR (lung tumour[Title/Abstract])) OR (pulmonary neoplasia[Title/Abstract])) OR (broncho-pulmonary cancer[Title/Abstract])) OR (bronchopulmonary cancer[Title/Abstract])) OR (NSCLC[Title/Abstract]))) AND (((((Radiomics[Title/Abstract]) OR (radiomic[Title/Abstract])) OR (radiogenomic[Title/Abstract])) OR (radiomics-based[Title/Abstract])) OR (Texture[Title/Abstract]))	1,844

2) Cochrane

Search number	Query	Results
#1	MeSH descriptor: [Lung Neoplasms] explode all trees	10639
#2	(Lung Neoplasms): ti,ab,kw OR (Pulmonary Neoplasms):ti,ab,kw OR (Lung Neoplasm):ti,ab,kw OR (Pulmonary Neoplasm):ti,ab,kw OR (Lung Cancer):ti,ab,kw	33016
#3	(Lung Cancers): ti,ab,kw OR (Pulmonary Cancer):ti,ab,kw OR (Lung Carcinoma):ti,ab,kw OR (Lung Carcinomas):ti,ab,kw OR (Pulmonary Cancer):ti,ab,kw	15870
#4	(Pulmonary Cancers): ti,ab,kw OR (Cancer of the Lung):ti,ab,kw OR (Cancer of Lung):ti,ab,kw OR (Adenocarcinoma of Lung):ti,ab,kw OR (Lung Adenocarcinomas):ti,ab,kw	30043
#5	(Lung Adenocarcinoma): ti,ab,kw OR (Small Cell Cancer Of The Lung):ti,ab,kw OR (Oat Cell Carcinoma of Lung):ti,ab,kw OR (lung tumor):ti,ab,kw OR (pulmonary tumour):ti,ab,kw	22716
#6	(pulmonary tumor): ti,ab,kw OR (broncho-pulmonary neoplasm):ti,ab,kw OR (broncho-pulmonary tumor):ti,ab,kw OR (broncho-pulmonary neoplasia):ti,ab,kw OR (bronchopulmonary neoplasm):ti,ab,kw	2040
#7	(bronchopulmonary tumor): ti,ab,kw OR (lung neoplasia):ti,ab,kw OR (lung tumour):ti,ab,kw OR (pulmonary neoplasia):ti,ab,kw OR (broncho-pulmonary cancer):ti,ab,kw	10672
#8	(bronchopulmonary cancer): ti,ab,kw OR (NSCLC):ti,ab,kw	11646
#9	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8	36575

#10	(Radiomics): ti,ab,kw OR (radiomic):ti,ab,kw OR (radiogenomic):ti,ab,kw OR (radiomics-based):ti,ab,kw OR (Texture):ti,ab,kw	2594
#11	#9 AND #10	153

3) Embase

Search number	Query	Results
#1	'lung tumor'/exp	564875
#2	'lung neoplasms':ab,ti OR 'pulmonary neoplasms':ab,ti OR 'lung neoplasm':ab,ti OR 'pulmonary neoplasm':ab,ti OR 'lung cancer':ab,ti OR 'lung cancers':ab,ti OR 'lung carcinoma':ab,ti OR 'lung carcinomas':ab,ti OR 'pulmonary cancer':ab,ti OR 'pulmonary cancers':ab,ti OR 'cancer of the lung':ab,ti OR 'cancer of lung':ab,ti OR 'adenocarcinoma of lung':ab,ti OR 'lung adenocarcinomas':ab,ti OR 'lung adenocarcinoma':ab,ti OR 'small cell cancer of the lung':ab,ti OR 'oat cell carcinoma of lung':ab,ti OR 'lung tumor':ab,ti OR 'pulmonary tumour':ab,ti OR 'pulmonary tumor':ab,ti OR 'broncho-pulmonary neoplasm':ab,ti OR 'broncho-pulmonary tumor':ab,ti OR 'bronchopulmonary neoplasia':ab,ti OR 'bronchopulmonary neoplasm':ab,ti OR 'bronchopulmonary tumor':ab,ti OR 'lung neoplasia':ab,ti OR 'lung tumour':ab,ti OR 'pulmonary neoplasia':ab,ti OR 'broncho-pulmonary cancer':ab,ti OR 'bronchopulmonary cancer':ab,ti OR nsclc:ab,ti	379080
#3	#1 OR #2	605675
#4	'radiomics'/exp	9517
#5	radiomics:ab,ti OR radiomic:ab,ti OR radiogenomic:ab,ti OR 'radiomics based':ab,ti OR texture:ab,ti	57004
#6	#4 OR #5	57897
#7	#3 AND #6	3175

4) Web of science

Search number	Query	Results
#1	Lung Neoplasms (subject) OR Pulmonary Neoplasms (subject) OR Lung Neoplasm (subject) OR Pulmonary Neoplasm (subject) OR Lung Cancer (subject) OR Lung Cancers (subject) OR Pulmonary Cancer (subject) OR Lung Carcinoma (subject) OR Lung Carcinomas (subject) OR Pulmonary Cancer (subject) OR Pulmonary Cancers (subject) OR Cancer of the Lung (subject) OR Cancer of Lung (subject) OR Adenocarcinoma of Lung (subject) OR Lung Adenocarcinomas (subject) OR Lung Adenocarcinoma (subject) OR Small Cell Cancer Of The Lung (subject) OR Oat Cell Carcinoma of Lung (subject) OR lung tumor (subject) OR pulmonary tumour (subject) OR pulmonary tumor (subject) OR broncho-pulmonary neoplasm (subject) OR broncho-pulmonary tumor (subject) OR bronchopulmonary neoplasia (subject) OR bronchopulmonary neoplasm (subject) OR bronchopulmonary tumor (subject) OR lung neoplasia (subject) OR lung tumour (subject) OR pulmonary neoplasia (subject) OR broncho-pulmonary cancer (subject) OR bronchopulmonary cancer (subject) OR NSCLC (subject)	553697
#2	Radiomics (subject) OR radiomic (subject) OR radiogenomic (subject) OR radiomics-based (subject) OR Texture (subject)	308059
#3	#1 AND #2	3385

Table S2.

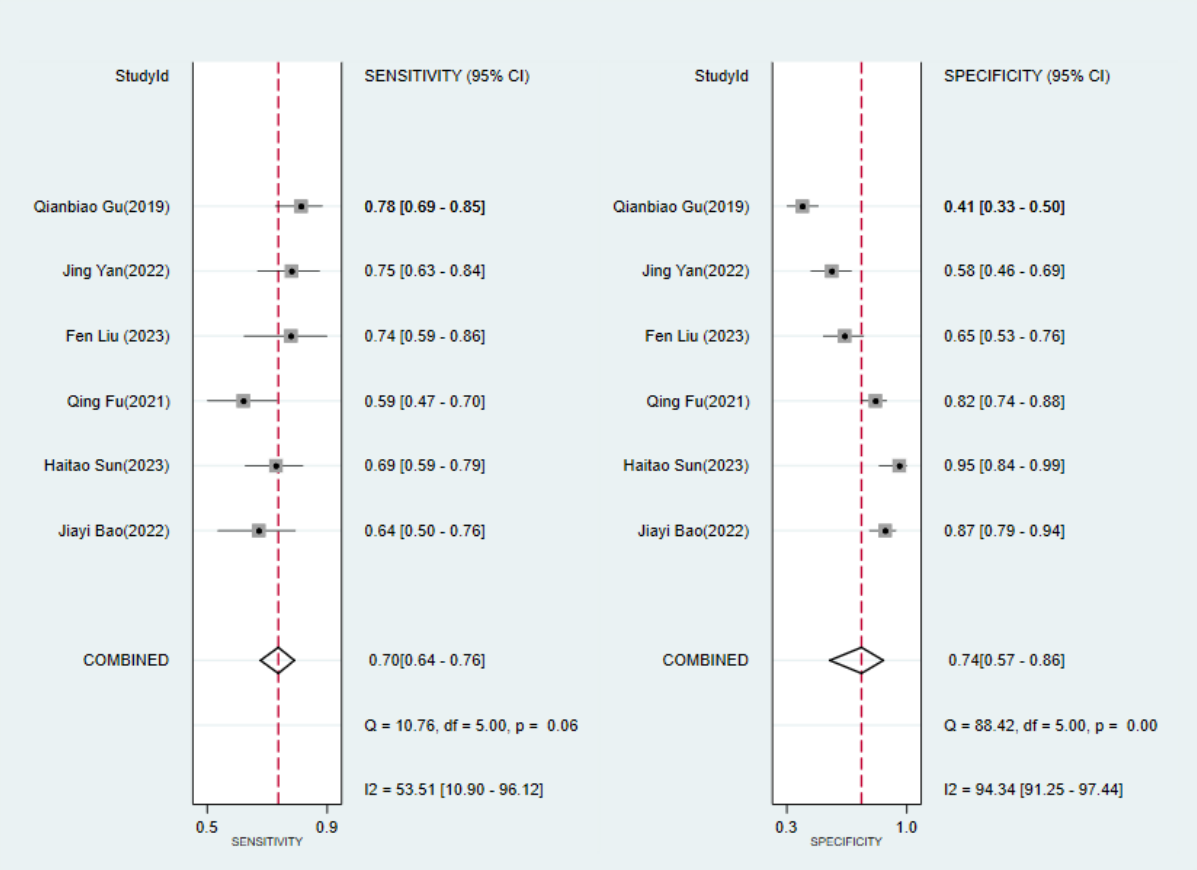
Number of cases of all outcome events	Total number	Number of cases in the	How to generate validation set	Number of cases in the	Model type

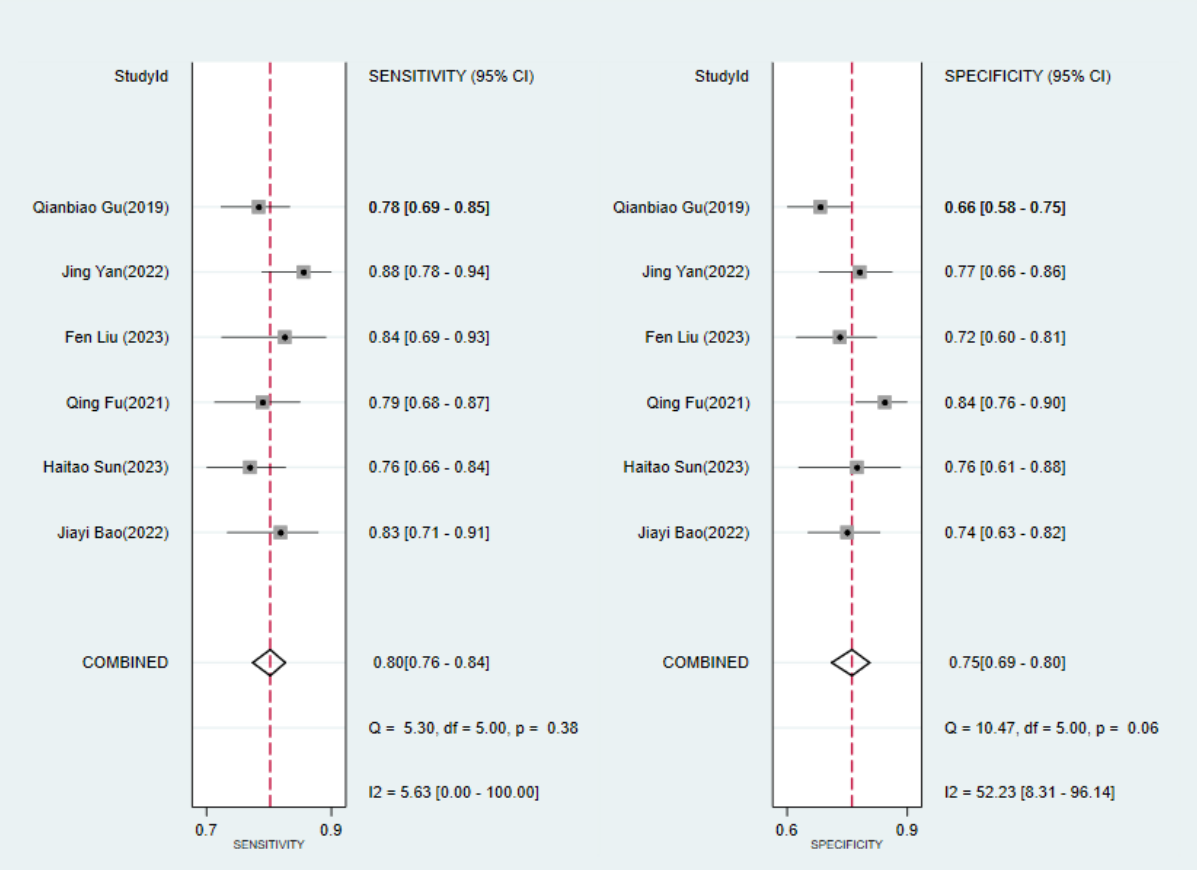
	of cases	training set		validation set	
42	137	95	Random sampling7:3	42	LR
Adenocarcinoma: 21, Squamous cell carcinoma:19	40	28	Random sampling7:3	12	Naı¨ve Bayes
	402	224	External validation	87 51	LR,SVM,AdaBo,RF,MLP,GB,as well asEn5(exceptGB)
Adenocarcinoma:51, Squamous cell carcinoma:147, Large cell carcinoma:111	309	247	Random sampling	62	RF, XGB, SVM, LR
Adenocarcinoma:86, Squamous cell carcinoma:85	171	171	10-fold cross validation		SVM
Adenocarcinoma:251, Squamous cell carcinoma:61, Small cell carcinoma:38, Large cell carcinoma:4	355	324		201	SVM,SMOTEfunction
Adenocarcinoma:47, Squamous cell carcinoma:40	87		Leave-one-out cross validation		SVM
Adenocarcinoma:240, Squamous cell carcinoma:110, Large cell carcinoma:108, others:59	317	317	10-fold cross validation		SVM
Adenocarcinoma:58, Squamous cell carcinoma:47	105	73	Random sampling	32	QDA, SVM with RBF kernel, SVM with sigmoid/tanh kernel, RF, and XGBoost
Adenocarcinoma:48, Squamous cell carcinoma:54	102	51	Five random samplings	51	CNN, LSTM+CNN, LSTM + CNN + SVM, LSTM + CNN + kNN
245	769	537	Random sampling	232	LR, DT, SVM, AB
Adenocarcinoma:324, Squamous cell carcinoma:110	434	434	10-fold cross validation		RBF-SVM,RF,KNN,RUSBoost
Adenocarcinoma:46,	71	48	Random	23	MR-Rad,CT-Rad,C

Squamous cell carcinoma:25			Sampling		T-Rad+MR-Rad
Adenocarcinoma:378, Squamous cell carcinoma:267	645	645	5-fold cross validation		LR, SVM, RF
86	153	75	Random sampling, external verification by institutions	78	LR
	1142	1002	Random sampling	140	LR,SVM,KNN,MLP,RF,XGBoost
Adenocarcinoma:705, Squamous cell carcinoma:583	1288	980	External validation	308	Bagging,AdaBoost,RF,XGBoost,GBDT,MLP,LR,GNB,SVM,KNN
Squamous cell carcinoma:152, Large cell carcinoma:106, Adenocarcinoma:150, uncategorized:58	466	373	Random sampling8:2	93	DA,KNN,SVM,NB,ensemble
Squamous cell carcinoma:121, Large cell carcinoma:101, Adenocarcinoma:71, uncategorized:56	349	349	10-fold cross validation		SLS
Adenocarcinoma:72, Squamous cell carcinoma:54	126	94	Random sampling	32	RF,LR,LR-L1,LR-PAC,CapsNet,CNN
80	211	117	Random sampling7:3, external validation	94	LR
Small Cell Lung Cancer:55, Adenocarcinoma:90, Squamous cell carcinoma:84	229	229	10-fold cross validation		Naïve Bayes, LR, RF
Adenocarcinoma: 55, Squamous cell carcinoma:66, Small Cell Lung Cancer:79	200	200	10-fold cross validation		SVM, LR, KNN, LDA, FNN
Squamous cell carcinoma: 134, Large cell carcinoma:110,	354				MRAS,Boruta

uncategorized: 62, Adenocarcinoma:48					
Adenocarcinoma:41, Squamous cell carcinoma:29	70	48	Random sampling7:3	22	LR
Adenocarcinoma:554, Squamous cell carcinoma:175, Small Cell Lung Cancer:191	920	644	Random sampling7:4	276	ProNet,com_radNet
117	245	245	10-fold cross validation		L2-LOG,LDA,CART,KNN,SVM,RF
107	282	197	Random sampling 7:3	85	LR
86	206	145	Random sampling7:3	61	LR
Adenocarcinoma: 87, Squamous cell carcinoma: 42	129	90	Random sampling7:3	39	LR
Adenocarcinoma:90, Squamous cell carcinoma:142, Adenosquamous carcinoma:81	313	246	External validation	67	GNB,RF,LR,SVM, GBM,XGBoost
52	132	87	External validation	45	RF
Adenocarcinoma:88, Squamous cell carcinoma:93, Small Cell Lung Cancer:97	278	278	3-fold cross validation		IFS, SVM

Supplementary Figures


StudyId
SENSITIVITY (95% CI)
Qianbiao Gu(2019) 0.78 [0.69 - 0.85]
Jing Yan(2022) 0.75 [0.63 - 0.84]
Fen Liu (2023) 0.74 [0.59 - 0.86]
Qing Fu(2021) 0.59 [0.47 - 0.70]
Haitao Sun(2023) 0.69 [0.59 - 0.79]
Jiayi Bao(2022) 0.64 [0.50 - 0.76]
COMBINED 0.70[0.64 - 0.76]
Q = 10.76, df = 5.00, p = 0.06
I2 = 53.51 [10.90 - 96.12]
0.5
0.9
SENSITIVITY
StudyId
SPECIFICITY (95% CI)
Qianbiao Gu(2019) 0.41 [0.33 - 0.50]
Jing Yan(2022) 0.58 [0.46 - 0.69]
Fen Liu (2023) 0.65 [0.53 - 0.76]
Qing Fu(2021) 0.82 [0.74 - 0.88]
Haitao Sun(2023) 0.95 [0.84 - 0.99]
Jiayi Bao(2022) 0.87 [0.79 - 0.94]
COMBINED 0.74[0.57 - 0.86]
Q = 88.42, df = 5.00, p = 0.00
I2 = 94.34 [91.25 - 97.44]
0.3
1.0
SPECIFICITY


StudyId
SENSITIVITY (95% CI)
Qianbiao Gu(2019) 0.78 [0.69 - 0.85]
Jing Yan(2022) 0.88 [0.78 - 0.94]
Fen Liu (2023) 0.84 [0.69 - 0.93]
Qing Fu(2021) 0.79 [0.68 - 0.87]
Haitao Sun(2023) 0.76 [0.66 - 0.84]
Jiayi Bao(2022) 0.83 [0.71 - 0.91]
COMBINED 0.80[0.76 - 0.84]
Q = 5.30, df = 5.00, p = 0.38
I2 = 5.63 [0.00 - 100.00]
0.7
0.9
SENSITIVITY
StudyId
SPECIFICITY (95% CI)
Qianbiao Gu(2019) 0.66 [0.58 - 0.75]
Jing Yan(2022) 0.77 [0.66 - 0.86]
Fen Liu (2023) 0.72 [0.60 - 0.81]
Qing Fu(2021) 0.84 [0.76 - 0.90]
Haitao Sun(2023) 0.76 [0.61 - 0.88]
Jiayi Bao(2022) 0.74 [0.63 - 0.82]
COMBINED 0.75[0.69 - 0.80]
Q = 10.47, df = 5.00, p = 0.06
I2 = 52.23 [8.31 - 96.14]
0.6
0.9
SPECIFICITY

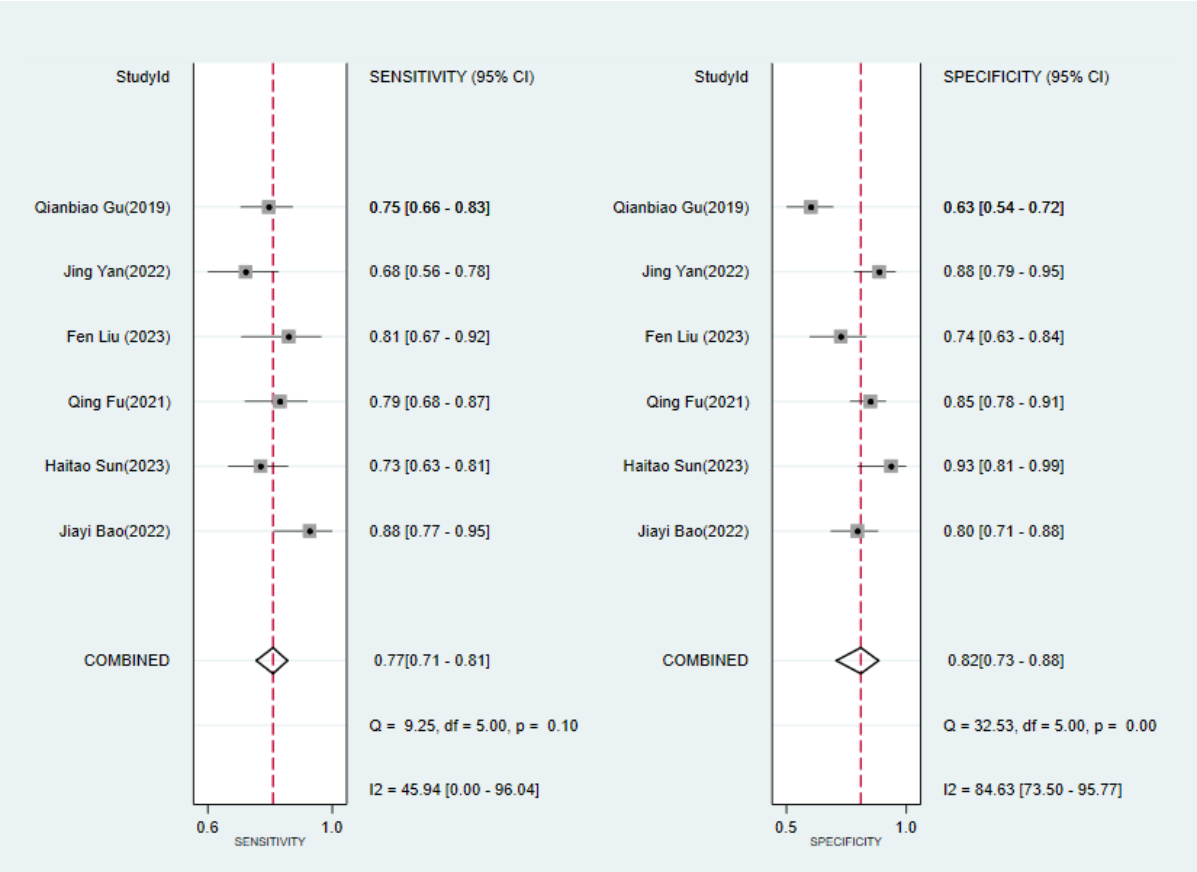
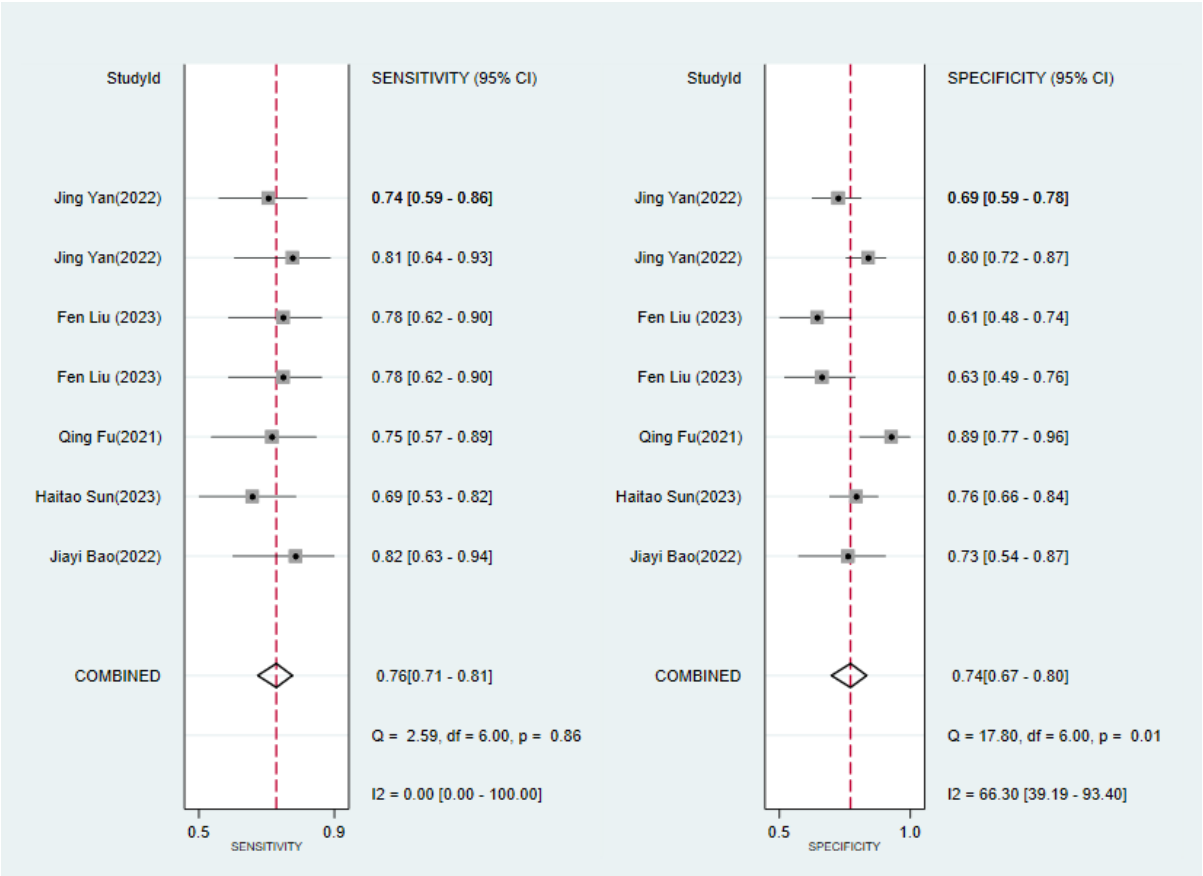
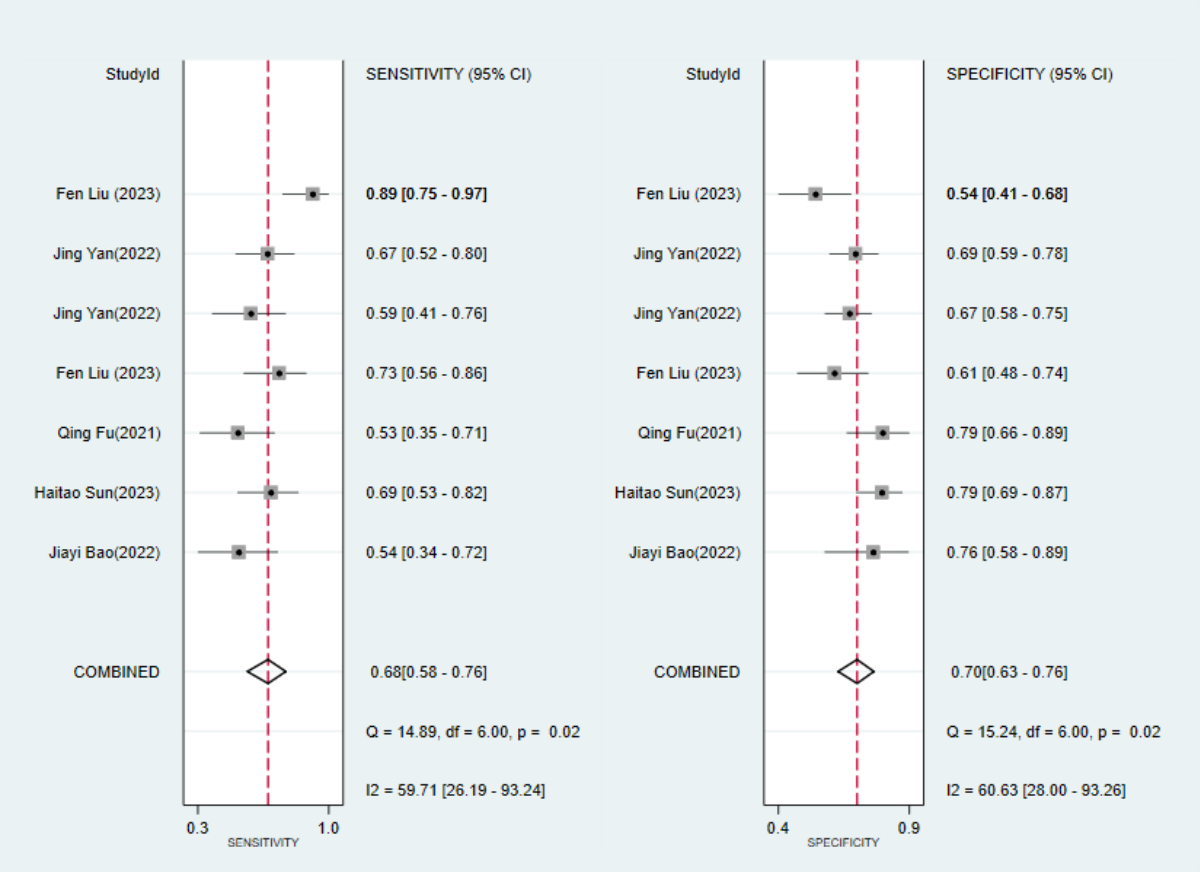


Figure S1-3. Sensitivity and specificity of radiomics and ML to identify Ki-67 expression in test sets





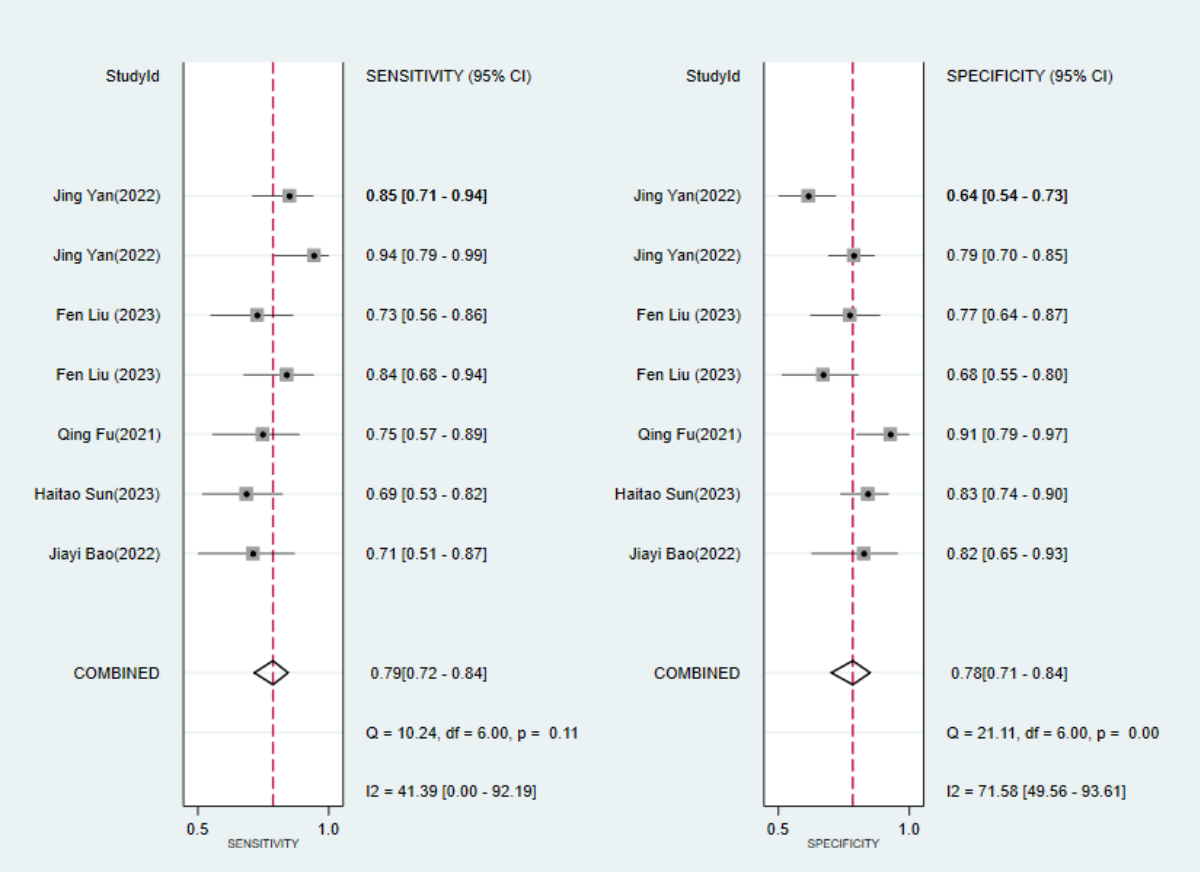


Figure S4-6. Sensitivity and specificity of radiomics and ML for identifying Ki-67 expression in validation sets

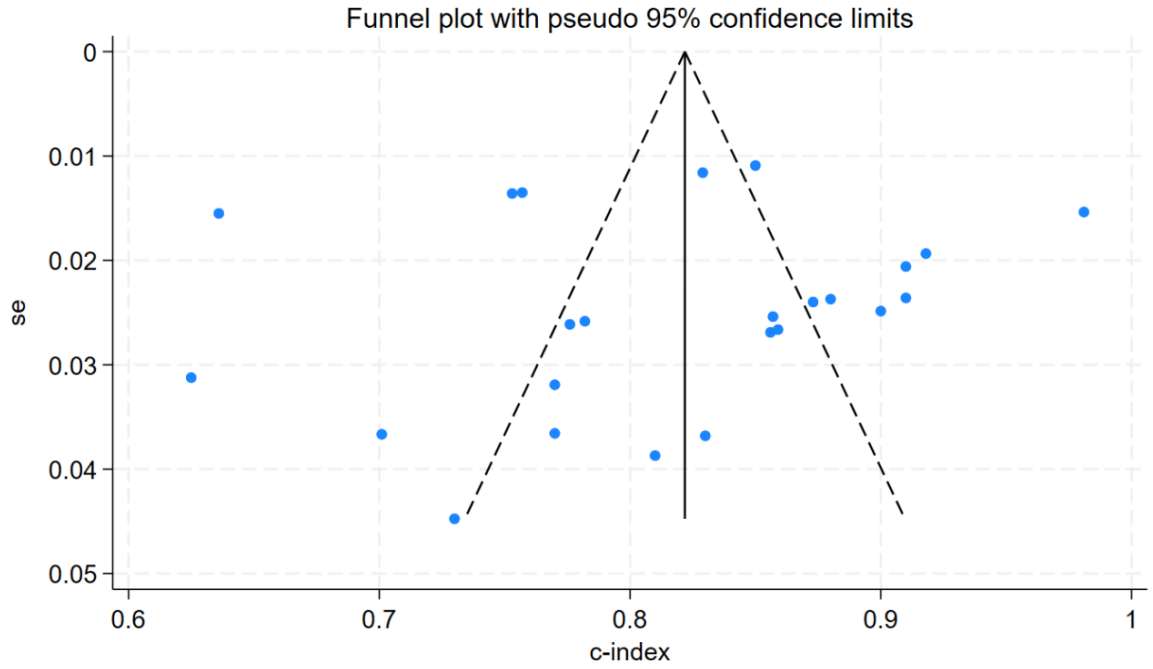


Figure S7. Funnel plot of meta-analysis covering imaging histology and ML to identify Ki-67 levels in the test sets

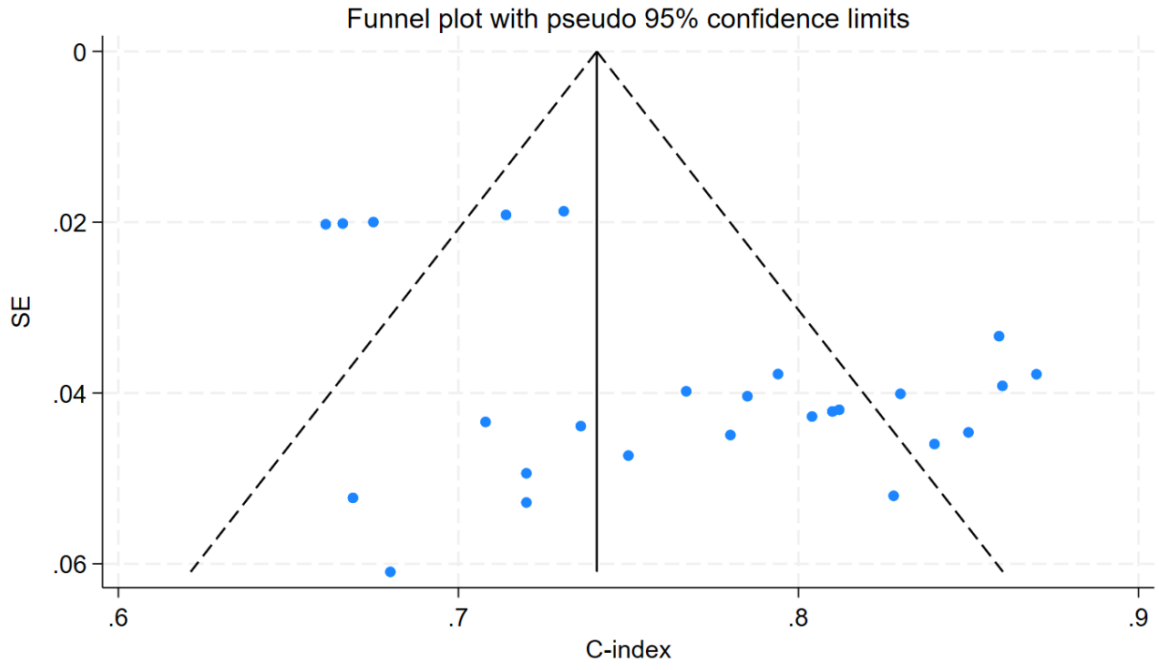


Figure S8. Funnel plot of meta-analysis covering imaging histology and ML to identify Ki-67 levels in the validation sets

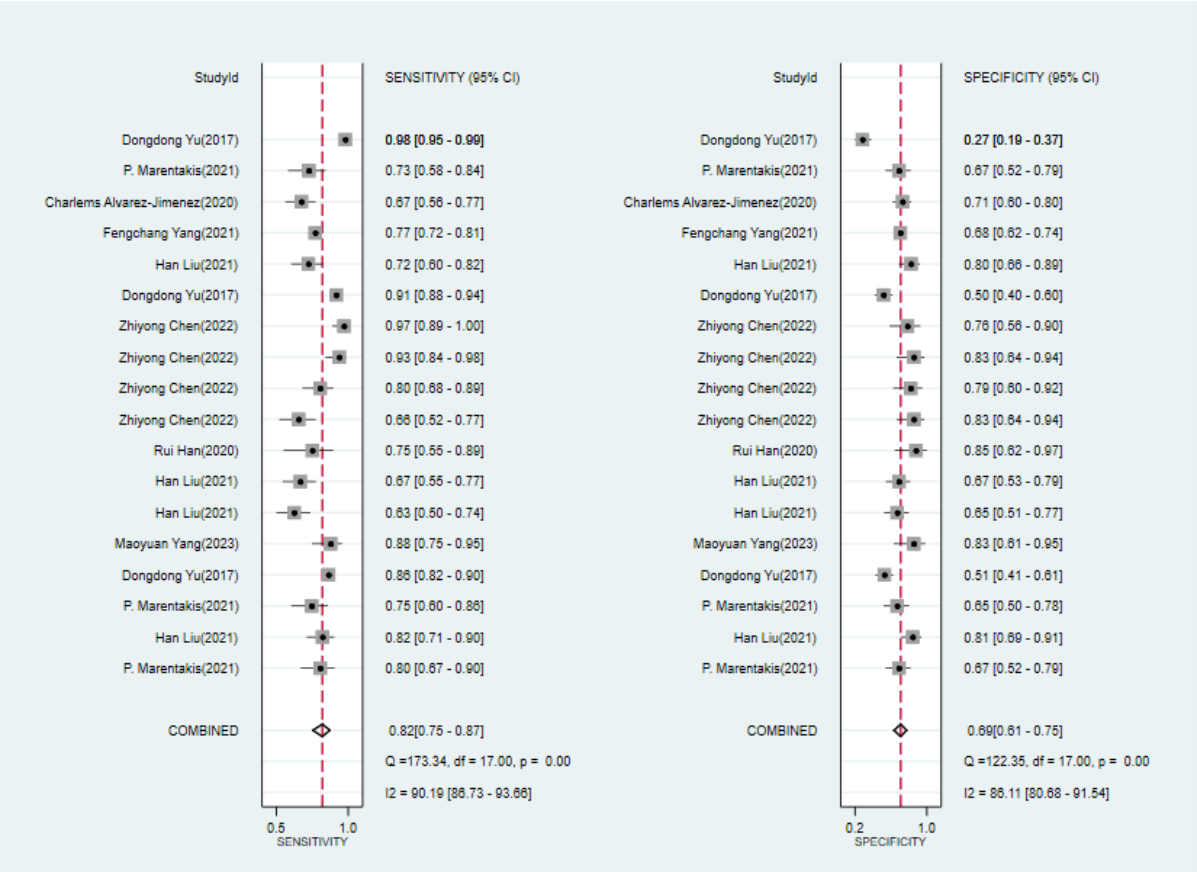


Figure S9. Sensitivity and specificity of radiomics in identifying different histological subtypes in the test sets

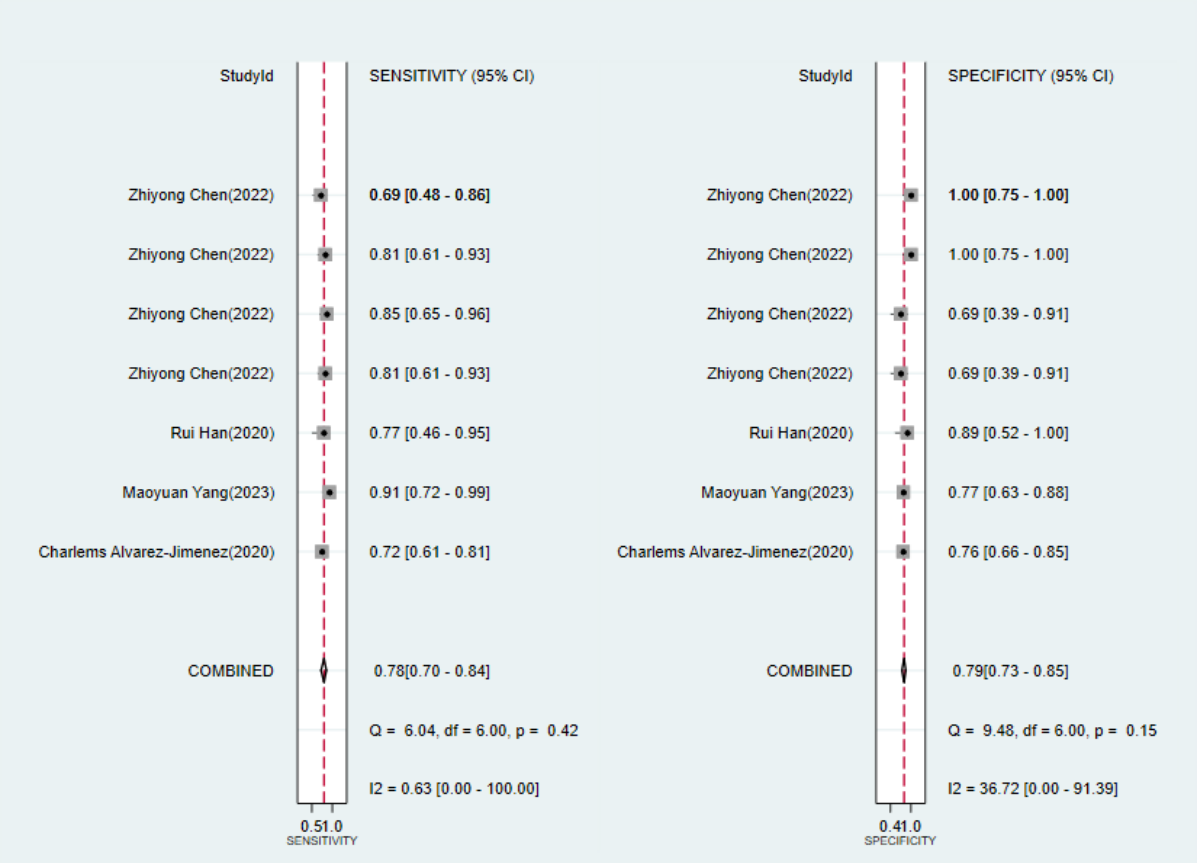


Figure S10. Sensitivity and specificity of radiomics in identifying different histological subtypes in the validation sets

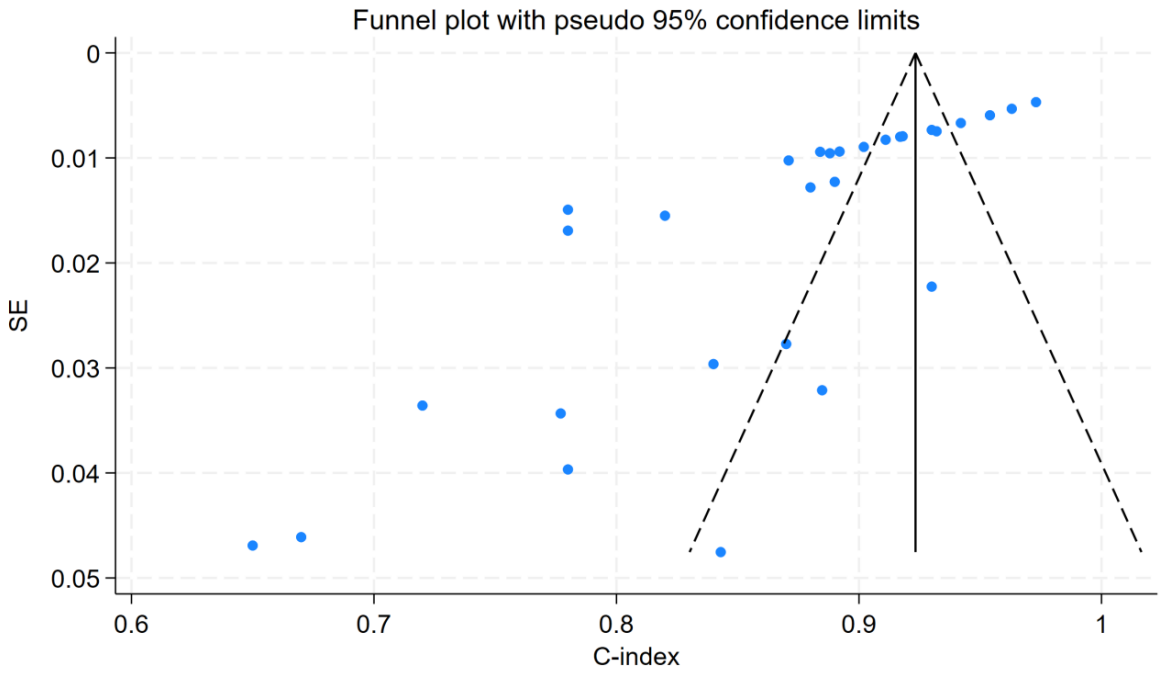


Figure S11. Funnel plot of meta-analysis of histological subtypes identified by radiomics in the test sets

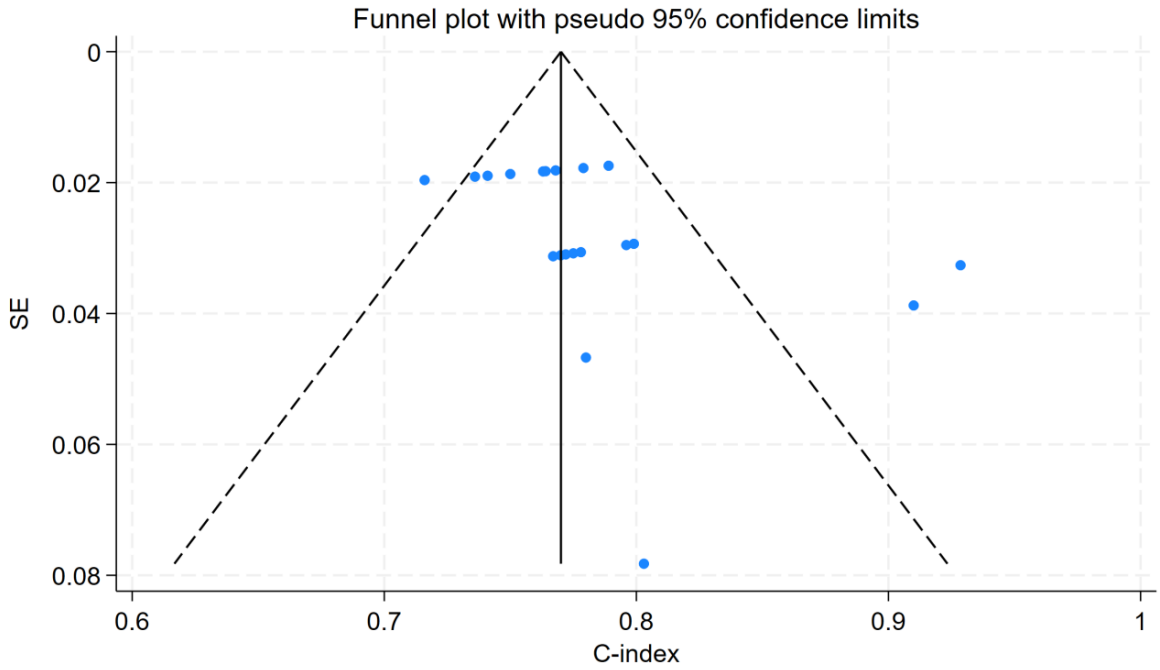


Figure S12. Funnel plot of meta-analysis of histological subtypes identified by radiomics in the validation sets