BMC Cancer

SYSTEMATIC REVIEW Open Access

Machine learning in prediction of epidermal growth factor receptor status in non-small cell lung cancer brain metastases: a systematic review and meta-analysis

Bardia Hajikarimloo[1*†], Ibrahim Mohammadzadeh[2†], Salem M. Tos[3], Mohammad Amin Habibi[4], Rana Hashemi[4], Ehsan Bahrami Hezaveh[4], Dorsa Najari[4], Arman Hasanzade[4], Mehdi Hooshmand[4] and Sara bana[4]

Abstract

Background Epidermal growth factor receptor (EGFR) mutations are present in 10–60% of all non-small cell lung cancer (NSCLC) patients and are associated with dismal prognosis. Lung cancer brain metastases (LCBM) are a common complication of lung cancer. Predictions of EGFR can help physicians in decision-making and, through optimizing treatment strategies, can result in more favorable outcomes. This systematic review and meta-analysis evaluated the predictive performance of machine learning (ML)-based models in EGFR status in NSCLC patients with brain metastasis.

Methods On December 20, 2024, the four electronic databases, Pubmed, Embase, Scopus, and Web of Science, were systematically searched. Studies that evaluated EGFR status in patients with brain metastasis from NSCLC were included.

Results Twenty studies with 3517 patients with 6205 NSCLC brain metastatic lesions were included. The majority of the best-performance models were ML-based (70%, 7/10), and deep learning (DL)-based models comprised 30% (6/20) of models. The area under the curve (AUC) and accuracy (ACC) of the best-performance models ranged from 0.765 to 1 and 0.69 to 0.93, respectively. The meta-analysis of the best-performance model revealed a pooled AUC of 0.91 (95%CI: 0.88–0.93) and ACC of 0.82 (95%CI: 0.79–0.86) along with a pooled sensitivity of 0.87 (95%CI: 0.83–0.9), specificity of 0.86 (95%CI: 0.79–0.9), and diagnostic odds ratio (DOR) of 35.2 (95%CI: 21.2–58.4). The subgroup analysis did not show significant differences between ML and DL models.

Conclusion ML-based models demonstrated promising predictive outcomes in predicting EGFR status. Applying ML-based models in daily clinical practice can optimize treatment strategies and enhance clinical and radiological outcomes.

†Bardia Hajikarimloo and Ibrahim Mohammadzadeh contributed equally to this work.

*Correspondence:
Bardia Hajikarimloo
bardii47@yahoo.com

Full list of author information is available at the end of the article

Keywords Machine learning, Deep learning, Epidermal growth factor receptor, EGFR, Lung neoplasms, Brain neoplasms

Introduction

Lung cancer (LC) stands as a principal cause of cancer-related death and is responsible for approximately 1.18 million annual deaths worldwide [1, 2]. Lung cancer brain metastases (LCBM) are a frequent complication in individuals with lung cancer, accounting for 10–36% of cases and comprising approximately 51% of all brain metastases [3, 4]. The incidence of LCBMs has escalated in recent years mainly due to advancements in imaging and prolonged survival of patients [5]. Small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), which comprises adenocarcinoma, squamous cell carcinoma, and large cell carcinoma, are the two main types of lung cancers (LCs) [6]. Tumor subtypes, molecular mutations, and patient attributes noticeably affect the development and progression of LCBM, necessitating personalized and tailored management [7]. The mounting incidence of LCBM underscores the necessity of enhancing diagnostic tools and treatment protocols to achieve more favorable outcomes.

Epidermal growth factor receptor (EGFR) mutations are identified in 10–60% of all NSCLC patients [8]. EGFR mutation mainly occurs at exon 19 and exon 21, and the 19 deletions and 21 L858R constitute about 85% of all mutation subtypes [9]. EGFR mutation is associated with a more dismal prognosis in LCBM patients as EGFR activation by ligand binding triggers receptor tyrosine kinase activity, resulting in cell proliferation and invasion [10, 11]. Recent investigations demonstrated that the application of tyrosine kinase inhibitors (TKIs) in LCBMs with EGFR mutation results in improved progression-free survival and radiological outcomes [12]. This indicates the importance of early detection of EGFR status in patients with LCBM when choosing the proper therapeutic intervention.

Substantial advancements in machine learning have recently led to the broad utilization of machine learning-based models in the medical field [13–16]. Several studies developed ML-based models to predict the EGFR status of NSCLC in patients with brain metastasis [17–36]. These models exhibited encouraging predictive performance in predicting EGFR status in NSCLC patients with brain metastasis [17–36]. Predictions of EGFR can help physicians in decision-making and, through optimizing treatment strategies, can result in more favorable clinical and radiological outcomes. This systematic review and meta-analysis evaluated the predictive performance of ML-based models in EGFR status in NSCLC patients with brain metastasis.

Materials and methods

Objective

This study aimed to assess the predictive performance of ML-based models for EGFR status in NSCLC patients with brain metastasis. It was based on the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)" guidelines [37].

Search strategy

On December 20, 2024, a comprehensive search was conducted in the four electronic databases: "PubMed", "Embase", "Scopus", and "Web of Science", using a unique search query for each database. The keywords were "Artificial intelligence," "Machine learning," "Deep learning," "Radiosurgery," "Epidermal growth factor receptor," "Lung Neoplasms,' and "Brain metastasis" and their equivalents. No limitations, such as "publication year" or "language", were used in the search process (Supplementary Table S1).

Eligibility criteria

To address the eligibility criteria, the following PICO was developed:

- Population (P): Patients with NSCLC brain metastasis.
- Intervention (I): Artificial intelligence (AI)-based models, including ML, deep learning (DL), and neural network (NN) developed to predict EGFR status.
- Comparison (C): None.
- Outcome (O): "Area under the curve (AUC)", "accuracy (ACC)", "sensitivity", and "specificity" for the prediction of EGFR status in LCBM.

The inclusion criteria were: (1) Studies that employed the ML-based models to predict EGFR status in NSCLC patients with brain metastasis and (2) Studies that reported "AUC", "ACC", "Sensitivity", and "Specificity" of the predictive model. The exclusion criteria were: (1) Inability to separate the predictive results for NSCLC brain metastasis from other LCBM types or brain metastases, (2) Lack of "AUC", "ACC", "Sensitivity", and "Specificity," of the model, (3) Book chapters, conference abstracts, preprints, commentaries, and editorials, and (4) Studies published in languages other than English.

Study selection process

The search results were imported into the "Covidence software". After identifying and resolving the duplicates,

two reviewers (B.H. and I.M.) screened the studies, and the third author (M.H.) resolved the conflicts. Studies that met the inclusion criteria were enrolled in the study.

Data extraction

Two authors performed the data using a predesigned Microsoft Excel datasheet. The data sheet comprised two sections: "Demographic characteristics" and "Predictive outcomes". The baseline section included the "Publication year", "Recruitment period", "Number of patients", "Number of lesions", "Mean age", "Number of males", "Number of females", "Number of smokers", "Number of non-smokers", "Number of EGFR mutation", "Number of EGFR wild-type", "Number of adenocarcinomas", "Number of non-adenocarcinoma", "Number of exon 18 mutation", "Number of exons 19 mutations", "Number of exons 20 mutations", "Number of exons 21 mutations". The performance section comprised the "ML model", "Algorithm", "Validation method", "Reference standard", "Outcome type", "AUC", "ACC", "Sensitivity", and "Specificity". Then, the data for the best-performance model were extracted based on the highest AUC or ACC for each OS, LC, and DR, if available.

Quality assessment

The "Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2)" tool was the tool for risk of bias (ROB) assessment [37]. The QUADAS-2 consists of four domains: "Patient selection", "Index test", "Reference standard", "Flow, and timing", each of which was answered by "High", "Low", or "Some concerns".

Statistical analysis

Statistical analyses were conducted in R language (R foundation of statistical computing V R-4.4.2). The "meta", "mada", "metafor", "dplyr", "officer", "flextable", and "ggplot" packages were used. The "True positive", "True negative", "False positive", and "False negative" values were calculated from "Sensitivity", "Specificity", "Prevalence", and "Sample size". The medians were converted to mean values through the method by Luo et al. [38]. An $I^2 > 50\%$ or Cochran's Q with $P < 0.05$ was considered substantial heterogeneity, and the random-effects model was used. The pooled "Sensitivity", "Specificity", "AUC", "ACC," and "Diagnostic odds ratio (DOR)" were calculated concurrently with their corresponding 95% confidence intervals (CIs). Pooled AUC and ACC estimates were calculated by the "metagen() function," using study-specific estimates by random-effects models with the Hartung-Knapp adjustment. Sensitivity and specificity were calculated by applying the"metaprop() function" with logit transformation and Clopper-Pearson confidence intervals. The "Hierarchical summary receiver operating characteristic (HSROC)" model was utilized to form a "Summary ROC (SROC)" curve. The robustness of the outcomes was evaluated through the leave-one-out sensitivity analysis.

Results

Study selection process

The comprehensive search of the literature resulted in the identification of 585 studies (Fig. 1). Of these, 234 were identified as duplicates and were omitted; as a result, 351 were enrolled for screening. During the title and abstract screening, 280 of the studied did not meet the eligibility criteria and were excluded. Therefore, 71 studies were included for full-text evaluation. During the dull-text screening, 43 studies were excluded due to irrelevancy, and eight were excluded due to conference abstracts. Ultimately, 20 studies were included for data extraction.

Quality assessment

The majority of the included studies exhibited a "Low" ROB across the four domains, especially in the "Index test" and "Fow and timing" (Supplementary Fig. S1-S2). However, some studies demonstrated "High" ROB in the "Patient selection" domain, resulting in concerns with recruitment protocols or eligibility criteria. The "Reference standard" and "Flow and timing" had an occasional evaluation of "Some concerns," but the overall ROB remained acceptable for most studies.

Baseline characteristics

Twenty studies with 3517 patients with 6205 NSCLC brain metastatic lesions were included (Table 1). The mean age ranged from 55.5 to 62.3 years. Female and male genders comprised 50.1% (1581/3155) and 49.9% (1574/3155) of the included population. Most included patients were non-smokers (60.2%, 1718/2854), while 39.8% (1136/2854) were smokers. Regarding the EGFR status, 54.8% (1930/3522) were EGFR mutated, and 45.2% (1592/3522) were EGFR wild-type. Adenocarcinoma was the primary histology in most available data (93.5%, 1066/1140). Of these 20 best-performance models, radiomics-based models (90%, 18/20) were the most frequent type of input features, followed by clinical-based (5%, 1/20) and clinical + radiomics-based (5%, 1/20).

Best-performance model characteristics

Table 2 illustrates the characteristics of the best-performance models. The majority of the best-performance models were ML-based (70%, 14/20) and DL-based models comprised 30% (6/20) of models. The most common validation method was train-test split (50%, 10/20), followed by 10-fold cross-validation (25%, 5/20), 5-fold cross-validation (20%, 4/20), and leave-one-out cross-validation (5%, 1/20). Least Absolute Shrinkage and Selection Operator (LASSO) (35%, 7/20) was the

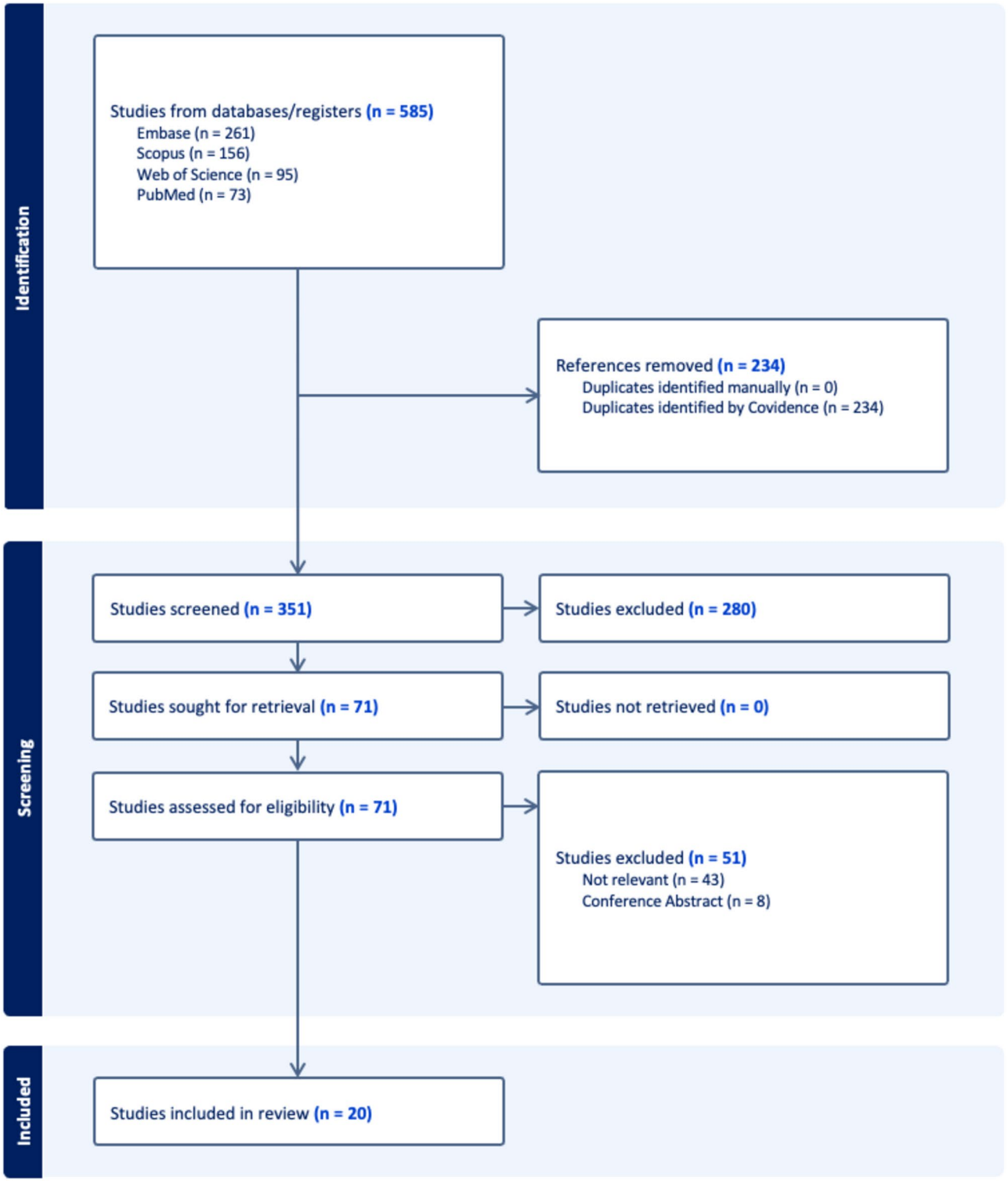


Fig. 1 PRISMA flowchart of the study selection process

most frequently used ML-based algorithm, followed by logistic regression (LR) (10%, 2/20), random forest (RF) (10%, 2/20), ResNet50 (10%, 2/20), and support vector machine (SVM) (10%, 2/20). The AUC and ACC of the best-performance models ranged from 0.765 to 1 and 0.69 to 0.93, respectively. The sensitivity and specificity of the best-performance models ranged from 0.687 to 1 and 0.55 to 1, respectively.

Table 1 Baseline characteristics of the included studies

Study	Year	Recruitment Period	Patients	Lesions	Mean age	Gender (M/F)	No. adenocarcinoma	No. non-adenocarcinoma	Smoker	EGFR status (M/W)	EGFR exons
Hu et al., 2015	2015	Jan 2003-Dec 2011	31	31	NA	12/19	22	9	13	12/19	0/8/0/4/0
Chen et al., 2020	2020	2009–2017	110	462	57.51	37/73	105	5	37	75/35	NA
Ahn et al., 2020	2020	Jun 2012-Jul 2018	61	210	NA	36/25	55	6	NA	29/32	0/14/11/3/NA
Park et al., 2021	2021	Aug 2011-Jun 2018	51	99	59.9	26/25	40	11	NA	42/57	NA/17/NA/6/NA
Zheng et al., 2022	2022	Jul 2014-Jul 2021	162	162	57	97/65	133	29	62	70/92	7/42/7/21/1
Shi et al., 2022	2022	Jan 2017-Sep 2020	60	60	57.82	30/30	NA	NA	NA	30/30	NA
Haim et al., 2022	2022	2009–2019	59	59	61.9	33/26	NA	NA	NA	16/43	NA
Fan et al., 2022 - Fusion	2022	Jul 2014-Oct 2021	310	310	NA	148/162	NA	NA	118	175/135	NA
Cao et al., 2022	2022	Jan 2017-Dec 2021	100	100	NA	57/43	100	0	44	49/51	NA
Zhou et al., 2023	2023	Aug 2018-Oct 2021	188	188	60.2	110/78	188	0	71	102/86	NA
Shi et al., 2023	2023	Nov 2017-Dec 2021	70	70	58.63	39/31	70	0	NA	35/35	NA
Mahajan et al., 2023	2023	2014–2018	117	117	NA	NA/NA	NA	NA	NA	33/41	NA
Fan et al., 2023 - Multiregional	2023	Jan 2014-Dec 2021	310	310	NA	NA/NA	NA	NA	NA	175/135	NA
Fan et al., 2023 - BTI	2023	Jan 2017-Dec 2021	310	310	NA	148/162	NA	NA	118	175/135	NA
Yang et al., 2024	2024	Jul 2014 -Dec 2022	293	293	NA	134/159	NA	NA	107	157/136	NA/74/NA/62/NA
Li et al., 2024	2024	Jan 2012- Jan 2023	399	2,646	NA	177/222	NA	NA	115	288/111	NA/150/NA/138/NA
Huang et al., 2024	2024	Jun 2014-Dec 2022	58	123	NA	77/46	58	0	45	31/27	1/13/4/9/NA
Hu et al., 2024	2024	Nov 2017-Dec 2021	359	186	NA	173/186	NA	NA	204	191/168	NA
Cao et al., 2024	2024	Jan 2018-Dec 2021	160	160	NA	78/82	NA	NA	85	84/76	0/41/0/43/0
Ouyang et al., 2025	2025	Jun 2017-Jun 2021	309	309	NA	162/147	295	14	117	161/148	NA

NA: Not available, M: Male, F: Female, LCBM: Lung cancer brain metastasis, EGFR = Epidermal growth factor receptor, M = Mutation, W = Wild-type

Table 2 Characteristics of all best-performance ML-based models

Study	ML model	Algorithm	Validation method	Reference standard	AUC	ACC	SEN	SPE
Hu et al., 2015	ML	SVM	5-fold CV	Clinical	0.9186	0.879	0.886	0.875
Chen et al., 2020	ML	RF	LOOCV	T1-CE, T2WI	0.912	0.777	0.731	0.906
Ahn et al., 2020	ML	SVM	10-fold CV	T1WI	0.8908	0.8906	0.8928	1
Park et al., 2021	ML	RF	10-fold CV	T1-CE, DTI	0.765	0.69	0.813	0.615
Zheng et al., 2022	ML	LASSO	Train-Test Split	T1-CE, T2WI, FLAIR	0.85	0.778	0.837	0.738
Shi et al., 2022	DL	ResNet50	Train-Test Split	T1-CE	0.85	0.7	1	0.55
Haim et al., 2022	DL	ResNet50	5-fold CV	T1-CE	0.91	0.898	0.687	0.977
Fan et al., 2022 - Fusion	ML	LASSO	10-fold CV	T1-CE, T2WI	0.945	NA	0.878	0.937
Cao et al., 2022	ML	LASSO	Train-Test Split	T1-CE, T2WI	0.968	0.93	0.902	0.959
Zhou et al., 2023	DL	MTSA-Net	Train-Test Split	T1-CE	NA	0.8788	0.8421	0.921
Shi et al., 2023	ML	LASSO	5-fold CV	T1-CE, T2WI	0.837	0.761	0.783	0.783
Mahajan et al., 2023	DL	EfficientNetB0	Train-Test Split	T1WI, T1-CE, T2WI, FLAIR	0.894	NA	NA	NA
Fan et al., 2023 - Multiregional	ML	LASSO	10-fold CV	T1-CE, T2WI	0.89	0.804	0.911	0.746
Fan et al., 2023 - BTI	ML	LASSO	10-fold CV	T1-CE, T2WI	0.896	0.817	0.822	0.841
Yang et al., 2024	ML	LASSO	Train-Test Split	T1-CE, T2WI, FLAIR	0.955	NA	0.891	0.897
Li et al., 2024	DL	Radio-GCN	Train-Test Split	DWI, T1-CE, T2WI, FLAIR	1	0.906	1	0.87
Huang et al., 2024	ML	LR	Train-Test Split	T1WI, T2WI, FLAIR, DWI, T1-CE, Clinical	0.814	0.767	0.889	0.634
Hu et al., 2024	ML	LR	Train-Test Split	T1-CE	0.885	0.825	0.914	0.758
Cao et al., 2024	DL	MSF-Net	5-fold CV	T1-CE, T2WI	0.96	0.81	0.9	0.87
Ouyang et al., 2025	ML	XGBoost	Train-Test Split	CT	0.889	0.863	0.752	0.677

NA: Not available, ML: Machine learning, DL: Deep learning, CV: Cross-validation, LOOCV: Leave-One-Out cross-validation, EGFR = Epidermal growth factor receptor, T1-CE: T1 contrast-enhanced, T2WI: T2-Weighted, FLAIR: Fluid-Attenuated Inversion Recovery, CT: Computed Tomography, SVM: Support Vector Machine, RF: Random Forest, LASSO: Least Absolute Shrinkage and Selection Operator, ResNet50: Residual Network 50, MTSA-Net: Multi-Scale Transformer-Aided Network, EfficientNetB0: Efficient Network Base Model 0, Radio-GCN: Radiomics-based Graph Convolutional Network, LR: Logistic Regression, MSF-Net: Multi-Scale Feature Network, XGBoost: Extreme Gradient Boosting, AUC: Area under the curve, ACC: Accuracy, SEN: Sensitivity, SPE: Specificity

Meta-analysis of outcomes

The meta-analysis of the best-performance model AUC revealed a pooled AUC of 0.91 (95%CI: 0.88–0.93) with substantial heterogeneity ($I^2 = 63\%$, $P < 0.001$) (Fig. 2). The subgroup analysis for AUC based on the ML model showed no significant difference between ML- and DL-based models (ML: 0.9 [95%CI: 0.88–0.93] vs. DL: 0.92 [95%CI: 0.85–0.99], $P = 0.54$) (Supplementary Fig. S3). The meta-analysis of the ACC demonstrated a pooled ACC of 0.82 (95%CI: 0.79–0.86) with high heterogeneity ($I^2 = 84.5\%$, $P < 0.001$) (Fig. 3). The subgroup analysis for ACC based on the ML model showed no significant difference between ML- and DL-based models (ML: 0.82 [95%CI: 0.78–0.86] vs. DL: 0.84 [95%CI: 0.73–0.95], $P = 0.68$) (Supplementary Fig. S4).

The meta-analysis of the sensitivity resulted in a pooled sensitivity of 0.87 (95%CI: 0.83–0.9) for best-performance model with high heterogeneity ($I^2 = 6.1\%$, $P < 0.001$) (Fig. 4). The subgroup analysis for sensitivity based on the ML model showed no significant difference between ML- and DL-based models (ML: 0.85 [95%CI: 0.81–0.89] vs. DL: 0.89 [95%CI: 0.84–0.92], $P = 0.28$) (Supplementary Fig. S5). The meta-analysis for specificity showed a pooled specificity of 0.86 (95%CI: 0.79–0.9) with high heterogeneity ($I^2 = 78.3\%$, $P < 0.001$) (Fig. 5). The subgroup analysis for specificity based on the ML model showed no significant difference between ML- and DL-based models (ML: 0.85 [95%CI: 0.76–0.91] vs. DL: 0.87 [95%CI: 0.74–0.95], $P = 0.67$) (Supplementary Fig. S6).

The meta-analysis for DOR exhibited a pooled DOR of 35.2 (95%CI: 21.2–58.4) with high heterogeneity ($I^2 = 71.5\%$, $P < 0.001$) (Fig. 5). The subgroup analysis for DOR based on the ML model showed no significant difference between ML- and DL-based models (ML: 28.9 [95%CI: 16.1–51.8] vs. DL: 66.8 [95%CI: 35.3-126.5], $P = 0.0057$) (Supplementary Fig. S7).

Summary receiver operating characteristic curve

The SROC curve demonstrated an AUC of 0.899 for forecasting EGFR status in NSCLC brain metastasis (Fig. 6). The estimated sensitivity is 0.851, with a false positive rate of 0.167, which indicates high sensitivity with a low false positive rate.

Sensitivity analysis

The leave-one-out sensitivity analysis of the AUC meta-analysis showed that the omission of each study did not significantly impact the AUC pooled estimated (Supplementary Fig. S8). Similarly, the leave-one-out sensitivity analysis of the ACC meta-analysis exhibited that the omission of each study did not significantly affect the AUC pooled estimated (Supplementary Fig. S9). The

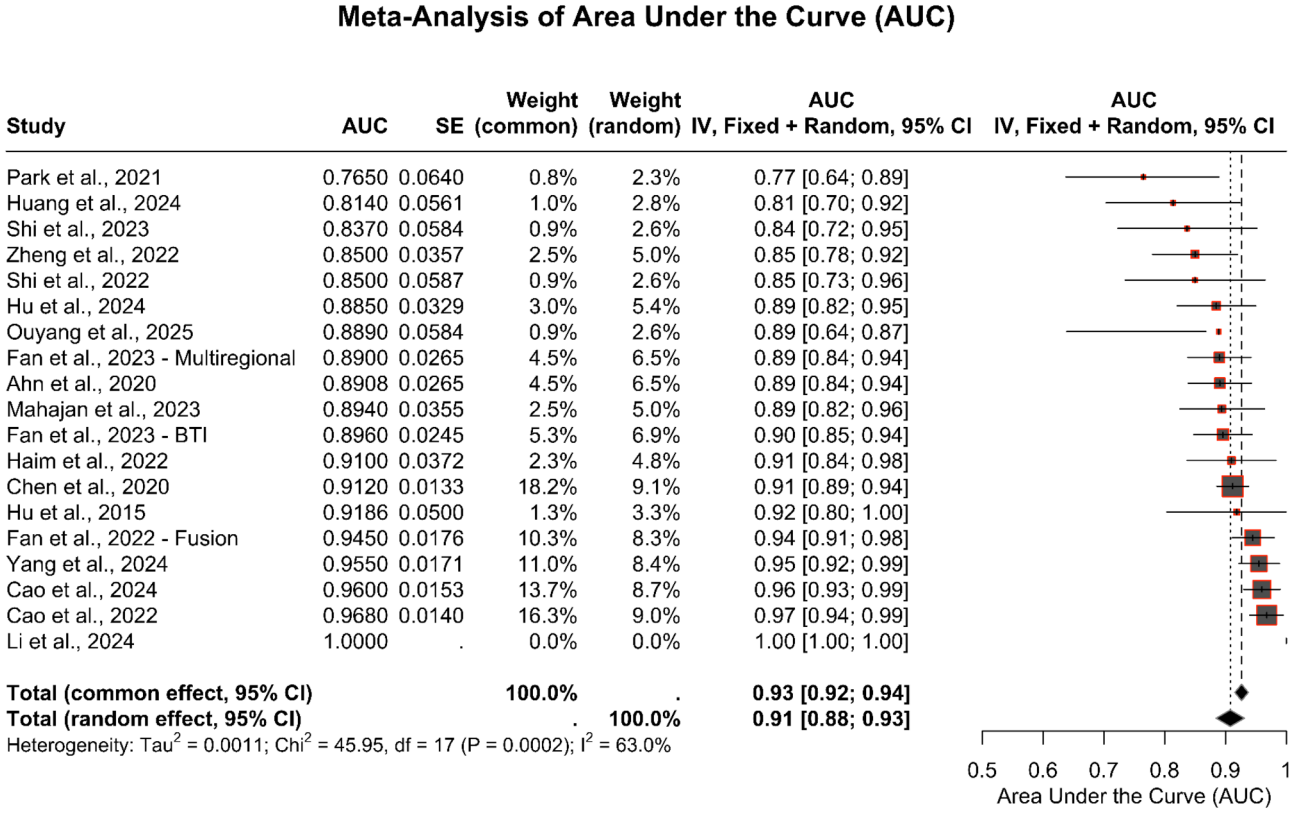


Fig. 2 Meta-analysis of the best-performance model area under the curve

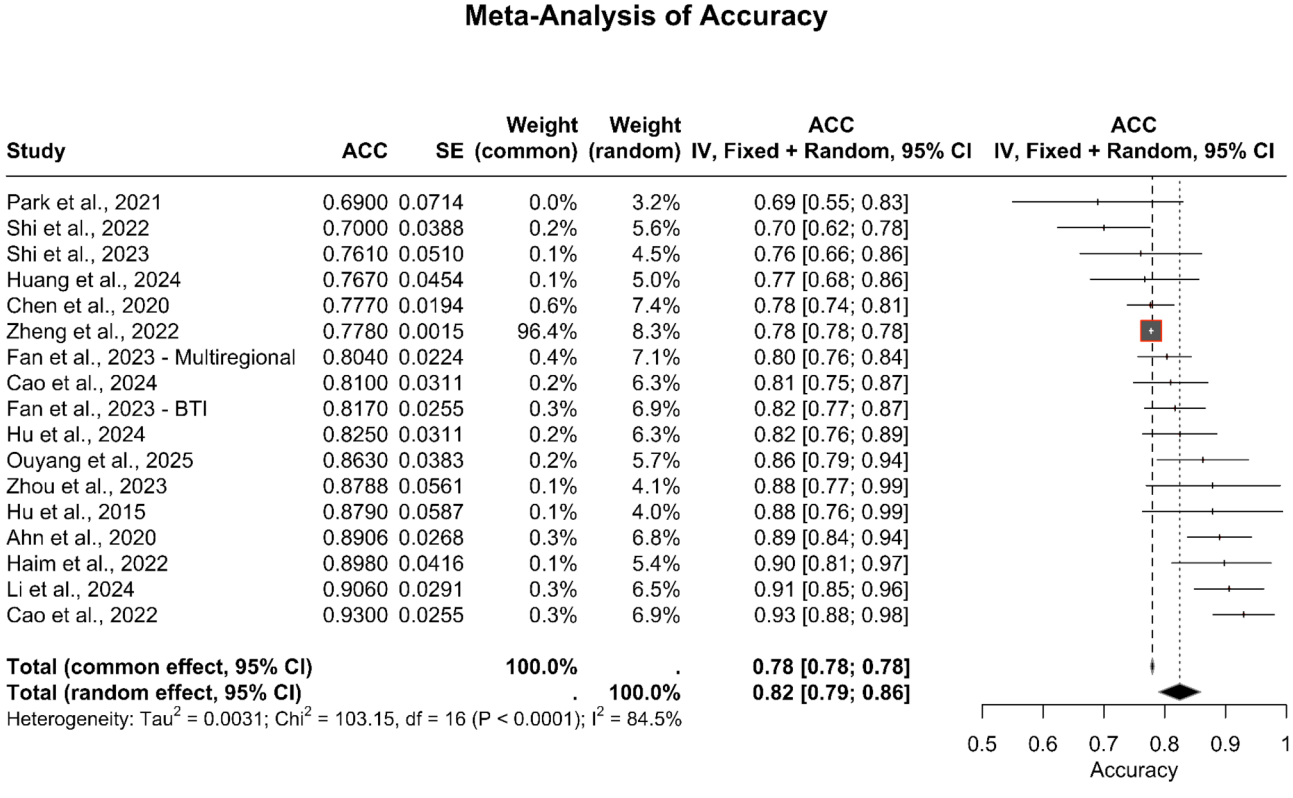


Fig. 3 Meta-analysis of the best-performance model accuracy

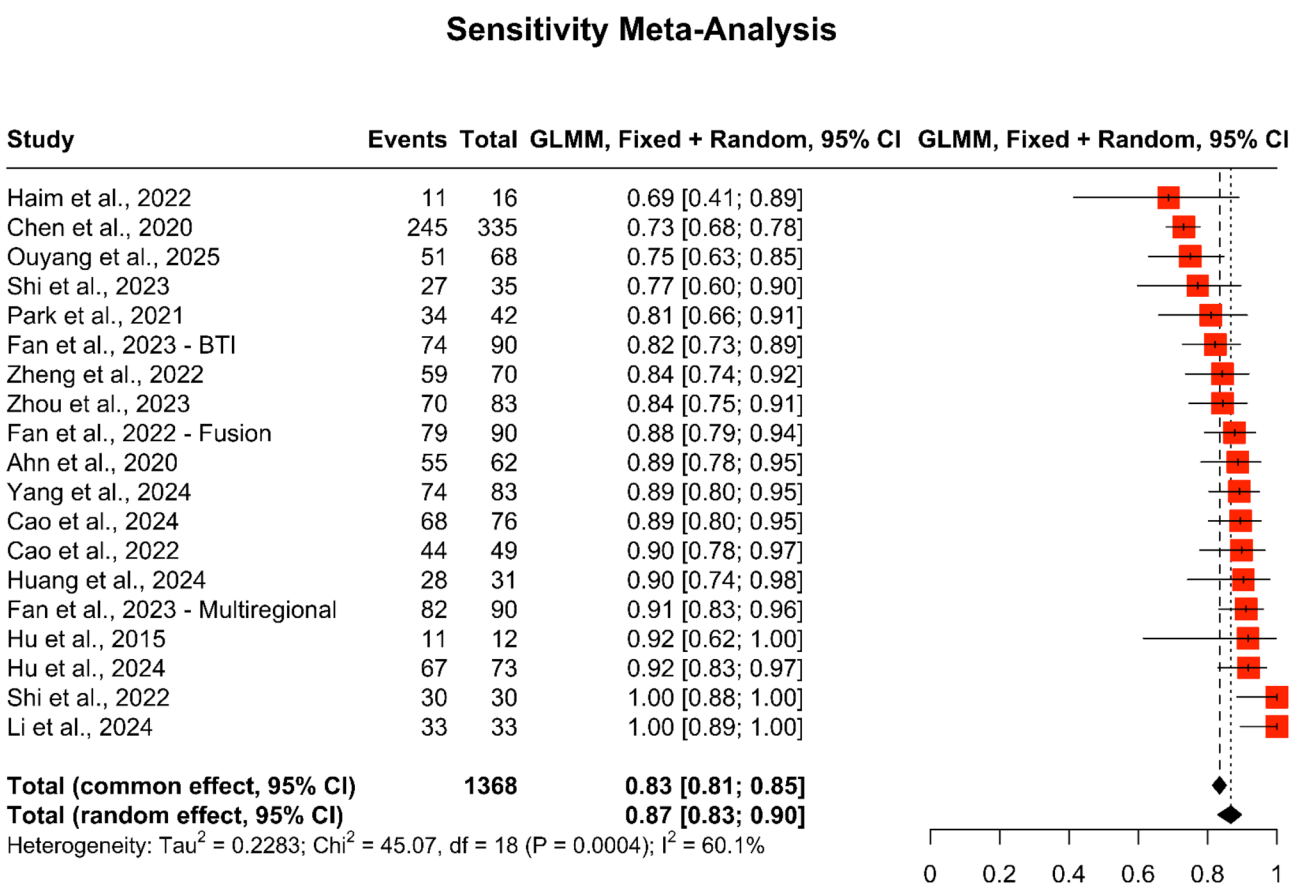


Fig. 4 Meta-analysis of the best-performance model sensitivity

leave-one-out sensitivity analysis of the sensitivity meta-analysis showed robust results, and the estimate consisted of (Supplementary Fig. S10). The leave-one-out sensitivity analysis of the specificity showed the high robustness of the results, and the pooled estimate remained stable (Supplementary Fig. S11). The leave-one-out sensitivity analysis of the DOR meta-analysis showed that the omission of each study did not significantly impact the DOR pooled estimate (Supplementary Fig. S12).

Discussion

Some mutations in the setting of LC lesions are associated with a higher likelihood of development of LCBM, including EGFR. Anaplastic lymphoma kinase (ALK), Kristen rat sarcoma viral oncogene homolog (KRAS), and ROS proto-oncogene 1 (ROS1) mutations [39]. EGFR mutation in LC tumors is associated with an increased risk of development of brain metastasis due to higher proliferation, invasiveness, and tendency to hematogenous spread than EGFR wild-type lesions [40, 41]. ML-based predictive models have revolutionized medical practice, especially for predicting molecular biomarkers involving EGFR status in NSCLC brain metastases [17–36]. These models utilize large-scale datasets to recognize sophisticated connections, providing unique potential for individualized medicine [42, 43]. Yet, the application of these models in daily clinical practice is challenging. Our systematic review and meta-analysis investigated the predictive performance of various ML and DL tools in forecasting EGFR status. Our findings showed that these models can achieve a pooled AUC of 0.91, ACC of 0.82, sensitivity of 0.87, specificity of 0.86, and DOR of 35.2 (Table 3).

ML models have several advantages, encompassing their substantial predictive performance [44]. ML-based models like SVM, RF, and LR achieve considerable accuracy by determining non-linear links in large-scale data [45]. For instance, the AUC of the included studies ranged from 0.765 to 1 [17–36]. In addition, ML-based tools are substantially adaptable and can incorporate high-dimensional data encompassing radiological, histopathological, and clinical parameters, enhancing prediction robustness [46, 47]. Another advantage of ML models is automation, which enables them to automate the decision-making process following training. Therefore, they can diminish the time needed for manual analysis and result in swift clinical employment [42, 48, 49]. Despite the encouraging advantages of ML-based models, these tools are

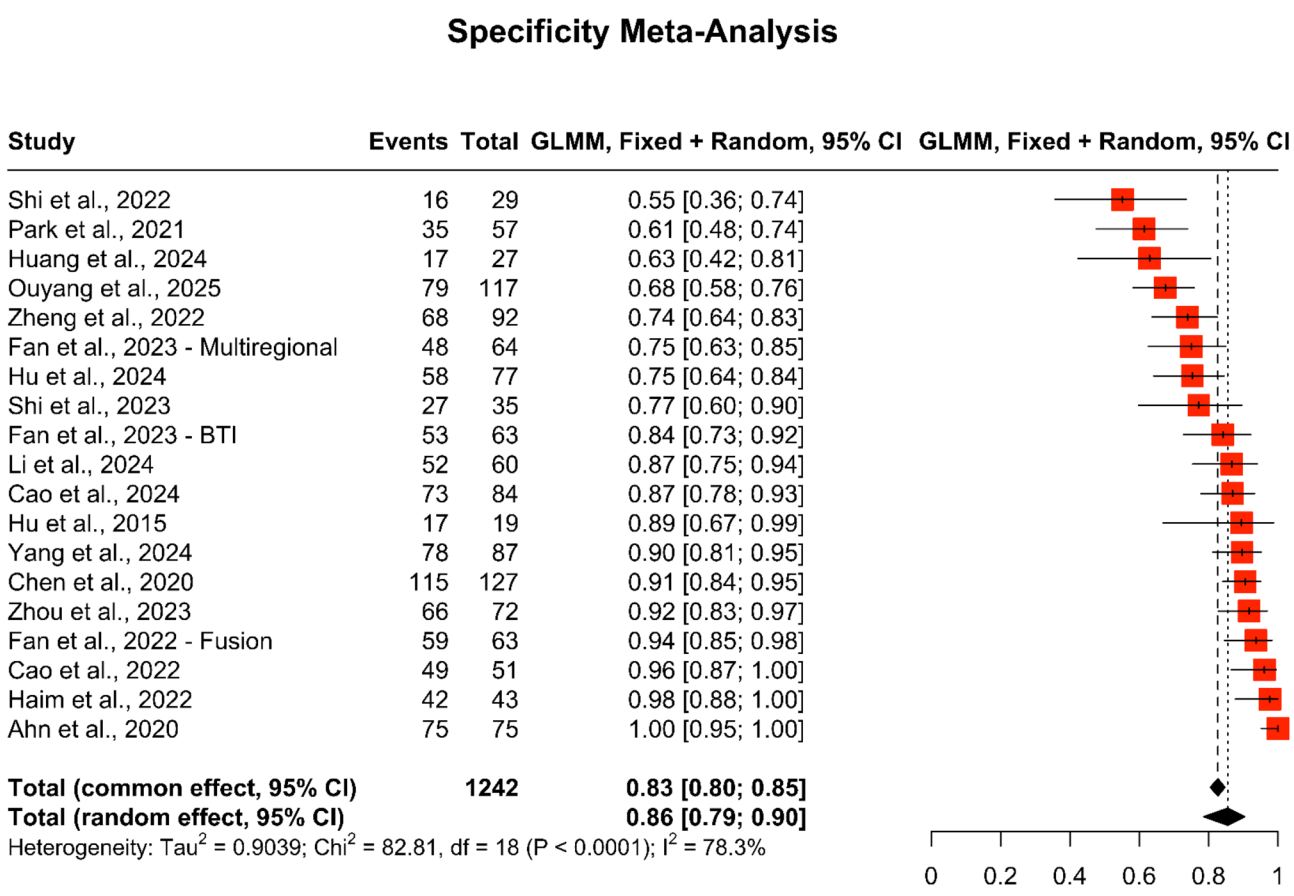
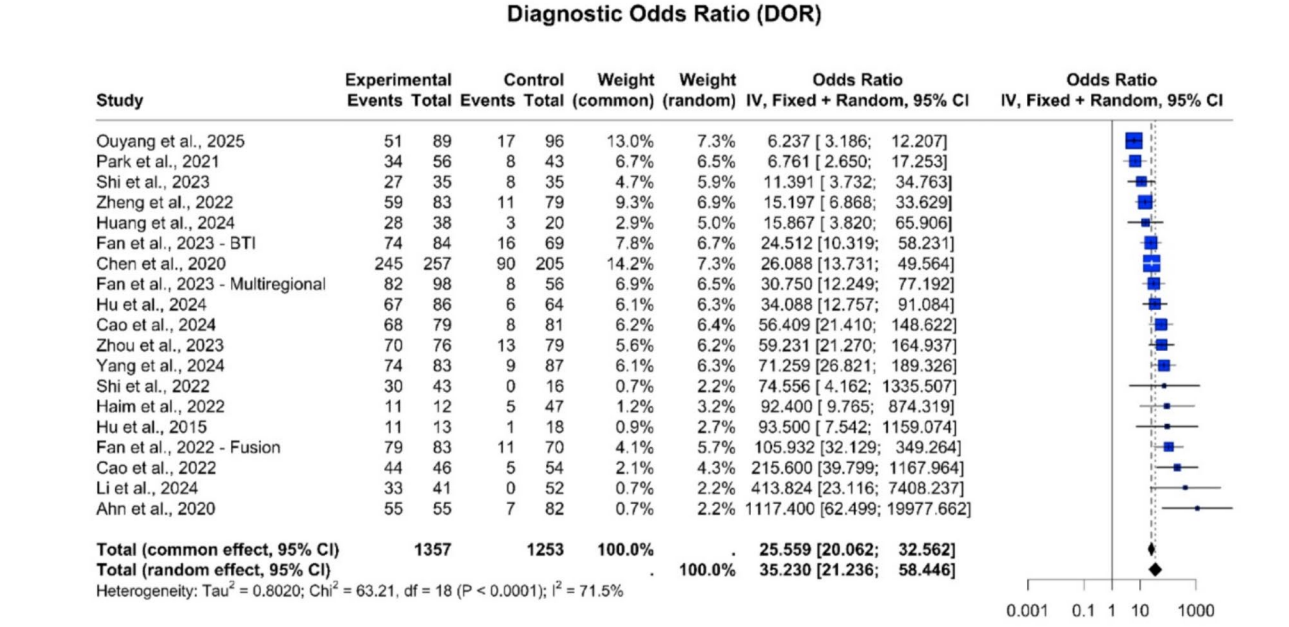


Fig. 5 Meta-analysis of the best-performance model specificity

associated with some limitations. ML models heavily rely on data quality and quantity, necessitating large, high-quality datasets for training and validation [50, 51]. A limited number of included participants can result in overfitting and poor generalizability [52, 53]. In addition, various developed ML models, especially DL-based models such as ResNet50 and Multi-Scale Fusion Network (MSF-Net), operate as "black boxes," resulting in challenging interpretability of these models that affect the explanation of them to patients by physicians [54–56].

ML-based models, unlike other predictive tools, provide distinct advantages and disadvantages, like nomograms. Traditional predictive tools are developed by applying predefined parameters and linear links to forecast results [57]. Traditional models' easy interpretability and ability to be applied by physicians in daily applications without advanced computational resources are the advantages of these models [58–60]. In contrast, these models are inherently restricted by the inability to determine advanced, non-linear relationships between the several variables [61]. ML models can analyze vast amounts of data and recognize the connection of various variables [14, 62].

Compared with DL tools, ML models demand fewer computational resources and provide easier interpretation and clinical utilization [63–65]. ML-based models are perfect for small datasets, providing interpretability and the capability to incorporate additional data, encompassing demographics and risk factors [63]. ML models apply hand-crafted radiomic characteristics that can be standardized and simply elucidated, yet they require substantial feature selection and data curation, extending the processing time [63]. Alternatively, DL models demonstrated exceptional capacity for large datasets, automatically capturing complex characteristics and handling advanced tasks, including radiological segmentation and classification [63]. By avoiding manual feature selection and performing transfer learning and data augmentation, DL models demand considerable computational resources, tend to overfit, and lack inherent interpretability, typically known as a "black box" [63]. ML-based models are generally preferable for smaller, structured datasets demanding interpretability, while DL-based tools are more suitable for large-scale, unstructured datasets requiring high performance and automation [63]. Our findings proposed that DL models have a slightly better predictive performance in forecasting EGFR status



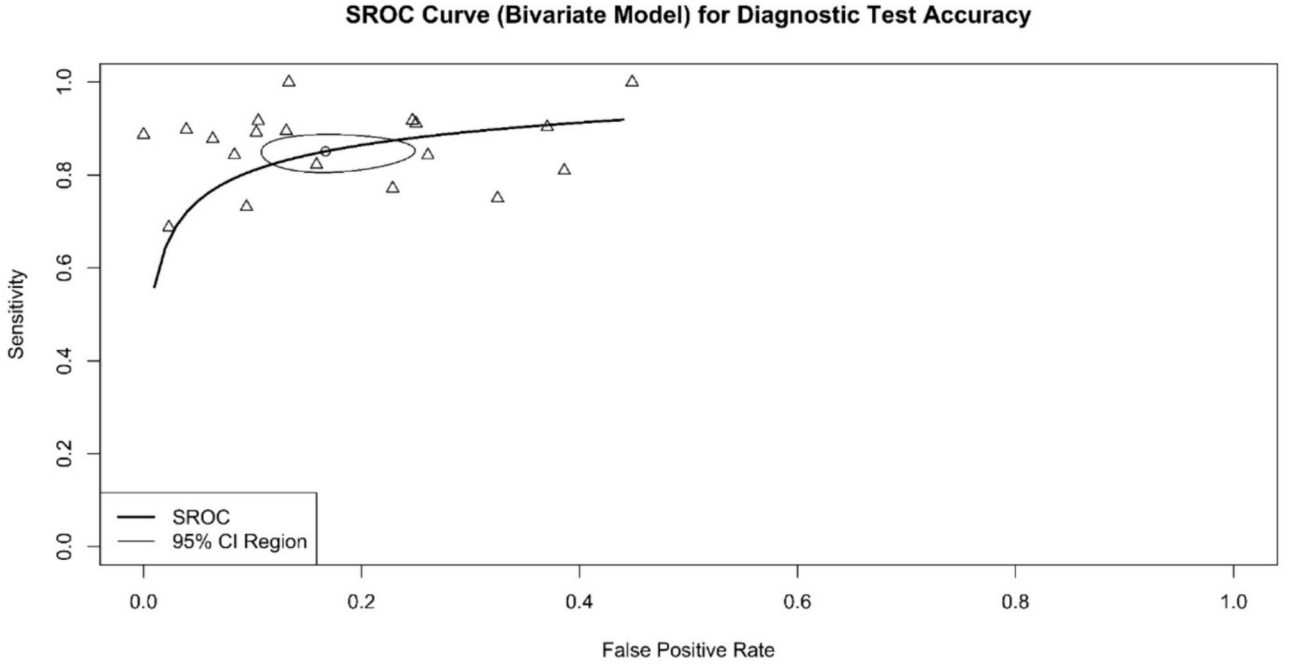


Fig. 6 SROC plot for ML-based models in the prediction of EGFR status in NSCLC brain metastasis

in NSCLC brain metastases; however, the superiority was insignificant.

Using ML models to forecast EGFR status in brain metastases from NSCLC can lead to groundbreaking upsides in clinical practice [17–36]. These models leverage imaging data, clinical characteristics, and complex imaging methods, like radiomics, to achieve precise and non-invasive predictions of EGFR mutation status [17–36]. These tools diminish the necessity for invasive tissue biopsies, reducing NSCLC patients' discomfort and intervention risks [17–36]. Moreover, by analyzing complex and high-dimensional data, ML models can recognize subtle patterns that traditional tools may miss, improving diagnostic accuracy [17–36]. Precise and early

Table 3 Summary of the meta-analysis results

Outcome	No. of studies	Overall estimate	ML	DL	*P*-value	Sensitivity Analysis
AUC	18	0.91 [95%CI: 0.88–0.93]	0.9 [95%CI: 0.88–0.93]	0.92 [95%CI: 0.85–0.99]	0.54	Robust
ACC	17	0.82 [95%CI: 0.79–0.86]	0.82 [95%CI: 0.78–0.86]	0.84 [95%CI: 0.73–0.95]	0.68	Robust
Sensitivity	19	0.87 [95%CI: 0.83–0.9]	0.85 [95%CI: 0.81–0.89]	0.89 [95%CI: 0.84–0.92]	0.28	Robust
Specificity	19	0.86 [95%CI: 0.79–0.9]	0.85 [95%CI: 0.76–0.91]	0.87 [95%CI: 0.74–0.95]	0.67	Robust
DOR	19	35.2 [95%CI: 21.2–58.4]	28.9 [95%CI: 16.1–51.8]	66.8 [95%CI: 35.3-126.5]	$P=0.0057$	Robust

No: Number, ML: Machine learning, DL: Deep learning, AUC: Area under the curve, ACC: Accuracy, DOR: Diagnostic odds ratio

prediction of EGFR status is crucial for the timely administration of targeted therapies, like TKIs, which are crucial in enhancing outcomes and quality of life in NSCLC individuals with brain metastases [17–36]. Integration of multimodal datasets and a combination of radiomics, clinical data, and histopathological biomarkers should be the main focus of further studies aiming to develop ML-based models in the prediction of EGFR in NSCLC patients. In addition, external validation should be conducted in independent cohorts to enhance the generalizability of the models and the facilitation of these models in clinical routine.

The heterogeneity seen in the results of the current meta-analysis arises from the variability in model performance, study designs, and datasets. Differences in sample size, individuals' baseline characteristics, radiological modalities, and distinctions in feature selection and validation methods increase the heterogeneity. Future models should utilize more standardized data and model assessment frameworks to diminish heterogeneity and increase the reliability of ML-based predictions.

Study limitations

Limited sample size in some of the included studies can affect the reliability and generalizability, along with concern about overfitting our findings' results. Most included models were developed based on retrospective data that may limit their application in real-world clinical practice. Most of the best-performance models were not based on external validation, which limits their applicability. Imbalance in the model type is another limitation, as ML models are more prominent, while DL models, which have the potential for cutting-edge performance, remain limited. However, we performed the subgroup analysis to overcome this limitation. Heterogeneity access to the models is another limitation of our study that may limit the generalizability of findings. Another limitation of our study is the predominance of ML models over DL models, which may limit the generalizability due to the requirement of larger datasets and a greater ability to capture more sophisticated patterns, which may result in enhanced function in real-world practice. This predominance of ML models can result in the application of simpler and more interpretable tools that may lead to the inability to fully use these AI models' potential. The lack of external validation in some of the models is another limitation of the study, which is associated with the likelihood of overfitting that can lead to mitigation of these models' reliability and clinical application. Further models with larger sample sizes and external validation are needed to verify our outcomes and their generalizability.

Conclusion

Early and accurate prediction of EGFR status is crucial in NSCLC patients with brain metastasis as it helps physicians in the decision-making process. ML-based models demonstrated promising predictive outcomes in predicting EGFR status; however, clinical validation is required to verify the application of ML tools in daily clinical practice. The application of ML-based models in daily clinical practice can optimize treatment strategies and enhance clinical and radiological outcomes. Further models with larger sample sizes and external validation are needed to verify our outcomes and their generalizability.

Abbreviations

SAH	Subarachnoid Hemorrhage
ML	Machine Learning
DL	Deep Learning
NN	Neural Networks
AUC	Area Under the Curve
ACC	Accuracy
DOR	Diagnostic Odds Ratio
GCS	Glasgow Coma Scale
IVH	Intraventricular Hemorrhage
ICH	Intracerebral Hemorrhage
VP	Ventriculoperitoneal
EVD	External Ventricular Drain
DRF	Distributed Random Forest
KNN	k-Nearest Neighbor
LASSO	Least Absolute Shrinkage and Selection Operator
RNN	Recurrent Neural Network
LR	Logistic Regression
CV	Cross-Validation
CSF	Cerebrospinal Fluid
GBM	Glioblastoma Multiforme
KPS	Karnofsky Performance Score
PFS	Progression-Free Survival

OS	Overall Survival
RF	Random Forest
SVM	Support Vector Machine
LOOCV	Leave-One-Out Cross-Validation
ADC	Apparent Diffusion Coefficient
T1-CE	Contrast-Enhanced T1-Weighted
T2-FLAIR	T2-Weighted Fluid-Attenuated Inversion Recovery
DW	Diffusion-Weighted
DTI	Diffusion Tensor Imaging
ANFIS	Adaptive Neuro-Fuzzy Inference System
DNN	Deep Neural Network
T1-W	T1-Weighted
EGFR	Epidermal Growth Factor Receptor
NSCLC	Non-Small Cell Lung Cancer
TKIs	Tyrosine Kinase Inhibitors

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-14221-w.

Supplementary Material 1

Acknowledgements
None.

Author contributions
Conceptualization: B.H. I.M., Methodology: B.H., M.H., Literature Search: B.H., Screening: I.M., R.H., Data Extraction: B.H., R.H., S.B., M.H., D.N., E.B., Risk of Bias Assessment: B.H., R.H., M.H., S.B., Statistical Analysis: B.H. E.B., Writing - Original Draft: B.H., S.T., M.H., I.M., Writing - Review & Editing: B.H., I.M., Supervision: B.H.

Funding
There is no funding source with authors to declare.

Data availability
The data supporting this study's findings are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The study is deemed exempt from receiving ethical approval.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Neurological Surgery, Shohada Tajrish Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2]Skull Base Research Center, Loghman-Hakim Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[3]Department of Neurological Surgery, University of Virginia, Charlottesville, VA, USA
[4]Department of Neurosurgery, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

Received: 23 January 2025 / Accepted: 24 April 2025
Published online: 01 May 2025

References

1. Wong MCS, Lao XQ, Ho KF, Goggins WB, Tse SLA. Incidence and mortality of lung cancer: global trends and association with socioeconomic status. Sci Rep [Internet]. 2017;7(1):14300. Available from: https://doi.org/10.1038/s41598-017-14513-7
2. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JWW, Comber H et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer [Internet]. 2013;49(6):1374–403. Available from: https://doi.org/10.1016/j.ejca.2012.12.027
3. Lamba N, Wen PY, Aizer AA. Epidemiology of brain metastases and leptomeningeal disease. Neuro Oncol [Internet]. 2021;23(9):1447–56. Available from: https://doi.org/10.1093/neuonc/noab101
4. Villano JL, Durbin EB, Normandeau C, Thakkar JP, Moirangthem V, Davis FG. Incidence of brain metastasis at initial presentation of lung cancer. Neuro Oncol [Internet]. 2015;17(1):122–8. Available from: https://doi.org/10.1093/neuonc/nou099
5. Al-Shamy G, Sawaya R. Management of brain metastases: the indispensable role of surgery. J Neurooncol [Internet]. 2009;92(3):275–82. Available from: https://doi.org/10.1007/s11060-009-9839-y
6. Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG. Introduction to the 2015 world health organization classification of tumors of the lung, pleura, thymus, and heart. J Thorac Oncol [Internet]. 2015;10(9):1240–2. Available from: https://doi.org/10.1097/JTO.0000000000000663
7. Ma DN, Gao XY, Dan YB, Zhang AN, Wang WJ, Yang G et al. Evaluating solid lung adenocarcinoma anaplastic lymphoma kinase gene rearrangement using noninvasive radiomics biomarkers. Onco Targets Ther [Internet]. 2020;13:6927–35. Available from: https://doi.org/10.2147/OTT.S257798
8. Midha A, Dearden S, McCormack R. EGFR mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutMapII). Am J Cancer Res [Internet]. 2015;5(9):2892–911. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26609494
9. Locatelli-Sanchez M, Couraud S, Arpin D, Riou R, Bringuier PP, Souquet PJ. Routine EGFR molecular analysis in non-small-cell lung cancer patients is feasible: exons 18–21 sequencing results of 753 patients and subsequent clinical outcomes. Lung [Internet]. 2013;191(5):491–9. Available from: https://doi.org/10.1007/s00408-013-9482-4
10. Selvaggi G, Novello S, Torri V, Leonardo E, De Giuli P, Borasio P et al. Epidermal growth factor receptor overexpression correlates with a poor prognosis in completely resected non-small-cell lung cancer. Ann Oncol [Internet]. 2004;15(1):28–32. Available from: https://doi.org/10.1093/annonc/mdh011
11. Levantini E, Maroni G, Del Re M, Tenen DG. EGFR signaling pathway as therapeutic target in human cancers. Semin Cancer Biol [Internet]. 2022;85:253–75. Available from: https://doi.org/10.1016/j.semcancer.2022.04.002
12. Rebuzzi SE, Alfieri R, La Monica S, Minari R, Petronini PG, Tiseo M. Combination of EGFR-TKIs and chemotherapy in advanced EGFR mutated NSCLC: Review of the literature and future perspectives. Crit Rev Oncol Hematol [Internet]. 2020;146(102820):102820. Available from: https://doi.org/10.1016/j.critrevonc.2019.102820
13. Raissi Dehkordi N, Raissi Dehkordi N, Karimi Toudeshki K, Farjoo MH. Artificial Intelligence in Diagnosis of Long QT Syndrome: A Review of Current State, Challenges, and Future Perspectives. Mayo Clinic Proceedings: Digital Health [Internet]. 2024;2(1):21–31. Available from: https://www.sciencedirect.com/science/article/pii/S2949761223000937
14. Hajikarimloo B, Habibi MA, Alvani MS, Meinagh SO, Kooshki A, Afkhami-Ardakani O et al. Machine learning-based models for prediction of survival in medulloblastoma: a systematic review and meta-analysis. Neurol Sci [Internet]. 2024;1–8. Available from: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=3C983WEAAAAJ&sortby=pubdate&citation_for_view=3C983WEAAAAJ:aqlVkmm33-oC
15. Hajikarimloo B, Sabbagh Alvani M, Koohfar A, Goudarzi E, Dehghan M, Hojjat SH et al. Clinical application of artificial intelligence in prediction of intraoperative cerebrospinal fluid leakage in pituitary surgery: A systematic review and meta-analysis. World Neurosurg [Internet]. 2024;191:303–13.e1. Available from: https://doi.org/10.1016/j.wneu.2024.09.015
16. Hajikarimloo B, Tos SM, Alvani MS, Rafiei MA, Akbarzadeh D, ShahirEftekhar M et al. Application of artificial intelligence in prediction of Ki-67 index in meningiomas: A systematic review and meta-analysis. World Neurosurg [Internet]. 2024; Available from: https://doi.org/10.1016/j.wneu.2024.10.089
17. Hu N, Wang G, Wu YH, Chen SF, Liu GD, Chen C et al. LDA-SVM-based EGFR mutation model for NSCLC brain metastases: an observational study. Medicine (Baltimore) [Internet]. 2015;94(5):e375. Available from: https://doi.org/10.1097/MD.0000000000000375
18. Chen BT, Jin T, Ye N, Mambetsariev I, Daniel E, Wang T et al. Radiomic prediction of mutation status based on MR imaging of lung cancer brain metastases. Magn Reson Imaging [Internet]. 2020;69:49–56. Available from: https://doi.org/10.1016/j.mri.2020.03.002

19. Ahn SJ, Kwon H, Yang JJ, Park M, Cha YJ, Suh SH et al. Contrast-enhanced T1-weighted image radiomics of brain metastases may predict EGFR mutation status in primary lung cancer. Sci Rep [Internet]. 2020;10(1):8905. Available from: https://doi.org/10.1038/s41598-020-65470-7
20. Park YW, An C, Lee J, Han K, Choi D, Ahn SS et al. Diffusion tensor and postcontrast T1-weighted imaging radiomics to differentiate the epidermal growth factor receptor mutation status of brain metastases from non-small cell lung cancer. Neuroradiology [Internet]. 2021;63(3):343–52. Available from: https://doi.org/10.1007/s00234-020-02529-2
21. Zheng L, Xie H, Luo X, Yang Y, Zhang Y, Li Y et al. Radiomic signatures for predicting EGFR mutation status in lung cancer brain metastases. Front Oncol [Internet]. 2022;12:931812. Available from: https://doi.org/10.3389/fonc.2022.931812
22. Shi J, Zhao Z, Jiang T, Ai H, Liu J, Chen X et al. A deep learning approach with subregion partition in MRI image analysis for metastatic brain tumor. Front Neuroinform [Internet]. 2022;16:973698. Available from: https://doi.org/10.3389/fninf.2022.973698
23. Haim O, Abramov S, Shofty B, Fanizzi C, DiMeco F, Avisdris N et al. Predicting EGFR mutation status by a deep learning approach in patients with non-small cell lung cancer brain metastases. J Neurooncol [Internet]. 2022;157(1):63–9. Available from: https://doi.org/10.1007/s11060-022-03946-4
24. Fan Y, Zhao Z, Wang X, Ai H, Yang C, Luo Y et al. Radiomics for prediction of response to EGFR-TKI based on metastasis/brain parenchyma (M/BP)-interface. Radiol Med [Internet]. 2022;127(12):1342–54. Available from: https://doi.org/10.1007/s11547-022-01569-3
25. Cao R, Pang Z, Wang X, Du Z, Chen H, Liu J et al. Radiomics evaluates the EGFR mutation status from the brain metastasis: a multi-center study. Phys Med Biol [Internet]. 2022;67(12):125003. Available from: https://doi.org/10.1088/1361-6560/ac7192
26. Zhou Z, Wang M, Zhao R, Shao Y, Xing L, Qiu Q et al. A multi-task deep learning model for EGFR genotyping prediction and GTV segmentation of brain metastasis. J Transl Med [Internet]. 2023;21(1):788. Available from: https://doi.org/10.1186/s12967-023-04681-8
27. Shi J, Chen H, Wang X, Cao R, Chen Y, Cheng Y et al. Using radiomics to differentiate brain metastases from lung cancer versus breast cancer, including predicting epidermal growth factor receptor and human epidermal growth factor receptor 2 status. J Comput Assist Tomogr [Internet]. 2023;47(6):924–33. Available from: https://doi.org/10.1097/RCT.0000000000001499
28. Mahajan A, B G, Wadhwa S, Agarwal U, Baid U, Talbar S et al. Deep learning based automated epidermal growth factor receptor and anaplastic lymphoma kinase status prediction of brain metastasis in non-small cell lung cancer. Explor Target Antitumor Ther [Internet]. 2023;4(4):657–68. Available from: https://doi.org/10.37349/etat.2023.00158
29. Fan Y, Wang X, Dong Y, Cui E, Wang H, Sun X et al. Multiregional radiomics of brain metastasis can predict response to EGFR-TKI in metastatic NSCLC. Eur Radiol [Internet]. 2023;33(11):7902–12. Available from: https://doi.org/10.1007/s00330-023-09709-7
30. Fan Y, Wang X, Yang C, Chen H, Wang H, Wang X et al. Brain-tumor interface-based MRI radiomics models to determine EGFR mutation, response to EGFR-TKI and T790M resistance mutation in non-small cell lung carcinoma brain metastasis. J Magn Reson Imaging [Internet]. 2023;58(6):1838–47. Available from: https://doi.org/10.1002/jmri.28751
31. Yang C, Fan Y, Zhao D, Wang Z, Wang X, Wang H et al. Habitat-based radiomics for predicting EGFR mutations in Exon 19 and 21 from brain metastasis. Acad Radiol [Internet]. 2024;31(9):3764–73. Available from: https://doi.org/10.1016/j.acra.2024.03.016
32. Li Y, Lv X, Chen C, Yu R, Wang B, Wang D et al. A deep learning model integrating multisequence MRI to predict EGFR mutation subtype in brain metastases from non-small cell lung cancer. Eur Radiol Exp [Internet]. 2024;8(1):2. Available from: https://doi.org/10.1186/s41747-023-00396-z
33. Huang Z, Tu X, Yu T, Zhan Z, Lin Q, Huang X. Peritumoural MRI radiomics signature of brain metastases can predict epidermal growth factor receptor mutation status in lung adenocarcinoma. Clin Radiol [Internet]. 2024;79(2):e305–16. Available from: https://doi.org/10.1016/j.crad.2023.10.022
34. Hu Y, Geng Y, Wang H, Chen H, Wang Z, Fu L et al. Improved prediction of epidermal growth factor receptor status by combined radiomics of primary nonsmall-cell lung cancer and distant metastasis. J Comput Assist Tomogr [Internet]. 2024;48(5):780–8. Available from: https://doi.org/10.1097/RCT.0000000000001591
35. Cao R, Fu L, Huang B, Liu Y, Wang X, Liu J et al. Brain metastasis magnetic resonance imaging-based deep learning for predicting epidermal growth factor receptor (EGFR) mutation and subtypes in metastatic non-small cell lung cancer. Quant Imaging Med Surg [Internet]. 2024;14(7):4749–62. Available from: https://doi.org/10.21037/qims-23-1744
36. Ouyang Z, Zhang G, He S, Huang Q, Zhang L, Duan X et al. CT and MRI bimodal radiomics for predicting EGFR status in NSCLC patients with brain metastases: A multicenter study. Eur J Radiol [Internet]. 2024;183(111853):111853. Available from: https://doi.org/10.1016/j.ejrad.2024.111853
37. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med [Internet]. 2011;155(8):529–36. Available from: https://doi.org/10.7326/0003-4819-155-8-201110180-00009
38. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Stat Methods Med Res [Internet]. 2018;27(6):1785–805. Available from: http://www.math.hkbu.edu.hk/~tongt/papers/SMMR2018.pdf
39. Kang Y, Jin Y, Li Q, Yuan X. Advances in lung cancer driver genes associated with brain metastasis. Front Oncol [Internet]. 2020;10:606300. Available from: https://doi.org/10.3389/fonc.2020.606300
40. Hochmair M. Medical treatment options for patients with epidermal growth factor receptor mutation-positive non-small cell lung cancer suffering from brain metastases and/or leptomeningeal disease. Target Oncol [Internet]. 2018;13(3):269–85. Available from: https://doi.org/10.1007/s11523-018-0566-1
41. Yousefi M, Bahrami T, Salmaninejad A, Nosrati R, Ghaffari P, Ghaffari SH. Lung cancer-associated brain metastasis: Molecular mechanisms and therapeutic options. Cell Oncol [Internet]. 2017;40(5):419–41. Available from: https://doi.org/10.1007/s13402-017-0345-5
42. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database [Internet]. 2020;2020. Available from: https://doi.org/10.1093/database/baaa010
43. Obijuru A, Arowoogun JO, Onwumere C, Odilibe IP, Anyanwu EC, Daraojimba AI. Big Data Analytics in healthcare: A review of recent advances and potential for Personalized Medicine. imsrj [Internet]. 2024;4(2):170–82. Available from: https://doi.org/10.51594/imsrj.v4i2.810
44. Alanazi HO, Abdullah AH, Qureshi KN. A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. J Med Syst [Internet]. 2017;41(4):69. Available from: https://doi.org/10.1007/s10916-017-0715-6
45. Guerrero MC, Parada JS, Espitia HE. EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks. Heliyon [Internet]. 2021;7(6):e07258. Available from: https://doi.org/10.1016/j.heliyon.2021.e07258
46. Xu X, Li J, Zhu Z, Zhao L, Wang H, Song C et al. A comprehensive review on synergy of multi-modal data and AI technologies in medical diagnosis. Bioengineering (Basel) [Internet]. 2024;11(3):219. Available from: https://doi.org/10.3390/bioengineering11030219
47. Zhou H, Zhou F, Zhao C, Xu Y, Luo L, Chen H. Multimodal data integration for precision oncology: Challenges and future directions [Internet]. arXiv [q-bio.QM]. 2024. Available from: http://arxiv.org/abs/2406.19611
48. Naugler C, Church DL. Automation and artificial intelligence in the clinical laboratory. Crit Rev Clin Lab Sci [Internet]. 2019;56(2):98–110. Available from: https://doi.org/10.1080/10408363.2018.1561640
49. Javaid M, Haleem A, Pratap Singh R, Suman R, Rab S. Significance of machine learning in healthcare: Features, pillars and applications. International Journal of Intelligent Networks [Internet]. 2022;3:58–73. Available from: https://doi.org/10.1016/j.ijin.2022.05.002
50. Hey T, Butler K, Jackson S, Thiyagalingam J. Machine learning and big scientific data. Philos Trans A Math Phys Eng Sci [Internet]. 2020;378(2166):20190054. Available from: https://doi.org/10.1098/rsta.2019.0054
51. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. J Big Data [Internet]. 2015;2(1). Available from: https://doi.org/10.1186/s40537-014-0007-7
52. Ghasemzadeh H, Hillman RE, Mehta DD. Toward generalizable machine learning models in speech, language, and hearing sciences: Estimating sample size and reducing overfitting. J Speech Lang Hear Res [Internet].

2024;67(3):753–81. Available from: https://doi.org/10.1044/2023_JSLHR-23-00273
53. Maleki F, Ovens K, Gupta R, Reinhold C, Spatz A, Forghani R. Generalizability of Machine Learning models: Quantitative evaluation of three methodological pitfalls. Radiol Artif Intell [Internet]. 2023;5(1):e220028. Available from: https://doi.org/10.1148/ryai.220028
54. Petch J, Di S, Nelson W. Opening the black box: The promise and limitations of explainable machine learning in cardiology. Can J Cardiol [Internet]. 2022;38(2):204–13. Available from: https://doi.org/10.1016/j.cjca.2021.09.004
55. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K et al. Interpreting black-box models: A review on explainable Artificial Intelligence. Cognit Comput [Internet]. 2024;16(1):45–74. Available from: https://doi.org/10.1007/s12559-023-10179-8
56. Zhang Z, written on behalf of AME Big-Data Clinical Trial Collaborative Group, Beck MW, Winkler DA, Huang B, Sibanda W et al. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. Ann Transl Med [Internet]. 2018;6(11):216–216. Available from: https://doi.org/10.21037/atm.2018.05.32
57. Staffa SJ, Zurakowski D. Statistical development and validation of clinical prediction models. Anesthesiology [Internet]. 2021;135(3):396–405. Available from: https://doi.org/10.1097/ALN.0000000000003871
58. Sun Y, Yang J, Li T, Gao K, Tong X. Nomogram for predicting facial nerve outcomes after surgical resection of vestibular schwannoma. Front Neurol [Internet]. 2021;12:817071. Available from: https://www.frontiersin.org/articles/https://doi.org/10.3389/fneur.2021.817071/full
59. Harris MK, Macielak RJ, Kaul VF, Zhang L, Moshtaghi O, Gao TZ et al. A nomogram to predict long-term facial nerve function after vestibular schwannoma resection: a contemporary multi-institutional study. J Neurosurg [Internet]. 2024;141(6):1667–74. Available from: https://doi.org/10.3171/2024.4.JNS232208
60. Shi X, Liu Y, Zhang Z, Tao B, Zhang D, Jiang Q et al. The value of radiographic features in predicting postoperative facial nerve function in vestibular schwannoma patients: A retrospective study and nomogram analysis. CNS Neurosci Ther [Internet]. 2024;30(4):e14526. Available from: https://doi.org/10.1111/cns.14526
61. Zheng H, Yan T, Han Y, Wang Q, Zhang G, Zhang L et al. Nomograms for prognostic risk assessment in glioblastoma multiforme: Applications and limitations. Clin Genet [Internet]. 2022;102(5):359–68. Available from: https://doi.org/10.1111/cge.14200
62. Guo C, Cui Y. Machine learning exhibited excellent advantages in the performance simulation and prediction of free water surface constructed wetlands. J Environ Manage [Internet]. 2022;309(114694):114694. Available from: https://doi.org/10.1016/j.jenvman.2022.114694
63. Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M et al. AI applications to medical images: From machine learning to deep learning. Phys Med [Internet]. 2021;83:9–24. Available from: https://doi.org/10.1016/j.ejmp.2021.02.006
64. Lai Y. A comparison of traditional machine learning and deep learning in image recognition. J Phys Conf Ser [Internet]. 2019;1314(1):012148. Available from: https://doi.org/10.1088/1742-6596/1314/1/012148
65. Korotcov A, Tkachenko V, Russo DP, Ekins S. Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. Mol Pharm [Internet]. 2017;14(12):4462–75. Available from: https://doi.org/10.1021/acs.molpharmaceut.7b00578

Publisher's note