Heliyon 10 (2024) e29605

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

The predictive value of serum tumor markers for EGFR mutation in non-small cell lung cancer patients with non-stage IA

Wenxing Du [a], Tong Qiu [a], Hanqun Liu [a], Ao Liu [a], Zhe Wu [a], Xiao Sun [a], Yi Qin [a], Wenhao Su [a], Zhangfeng Huang [a], Tianxiang Yun [b], Wenjie Jiao [a,*]

[a] Department of Thoracic Surgery, Affiliated Hospital of Qingdao University, Qingdao, China
[b] Department of Thoracic Surgery, The Second Affiliated Hospital, Shandong First Medical University, Taian, China

ARTICLE INFO

Keywords:
Lung cancer
Epidermal growth factor receptor
Serum tumor markers
Nomogram model
Machine learning

ABSTRACT

Objective: The predictive value of serum tumor markers (STMs) in assessing epidermal growth factor receptor (EGFR) mutations among patients with non-small cell lung cancer (NSCLC), particularly those with non-stage IA, remains poorly understood. The objective of this study is to construct a predictive model comprising STMs and additional clinical characteristics, aiming to achieve precise prediction of EGFR mutations through noninvasive means.

Materials and methods: We retrospectively collected 6711 NSCLC patients who underwent EGFR gene testing. Ultimately, 3221 stage IA patients and 1442 non-stage IA patients were analyzed to evaluate the potential predictive value of several clinical characteristics and STMs for EGFR mutations.

Results: EGFR mutations were detected in 3866 patients (57.9 %) of all NSCLC patients. None of the STMs emerged as significant predictor for predicting EGFR mutations in stage IA patients. Patients with non-stage IA were divided into the study group (n = 1043) and validation group (n = 399). In the study group, univariate analysis revealed significant associations between EGFR mutations and the STMs (carcinoembryonic antigen (CEA), squamous cell carcinoma antigen (SCC), and cytokeratin-19 fragment (CYFRA21-1)). The nomogram incorporating CEA, CYFRA 21-1, pathology, gender, and smoking history for predicting EGFR mutations with non-stage IA was constructed using the results of multivariate analysis. The area under the curve (AUC = 0.780) and decision curve analysis demonstrated favorable predictive performance and clinical utility of nomogram. Additionally, the Random Forest model also demonstrated the highest average C-index of 0.793 among the eight machine learning algorithms, showcasing superior predictive efficiency.

Conclusion: CYFRA21-1 and CEA have been identified as crucial factors for predicting EGFR mutations in non-stage IA NSCLC patients. The nomogram and 8 machine learning models that combined STMs with other clinical factors could effectively predict the probability of EGFR mutations.

* Corresponding author. Department of Thoracic Surgery, Affiliated Hospital of Qingdao University, NO.16 Jiangsu road, Qingdao, Shandong Province, 266071, China.
E-mail address: jiaowj@qduhospital.cn (W. Jiao).

https://doi.org/10.1016/j.heliyon.2024.e29605
Received 1 February 2024; Received in revised form 7 April 2024; Accepted 10 April 2024
Available online 15 April 2024

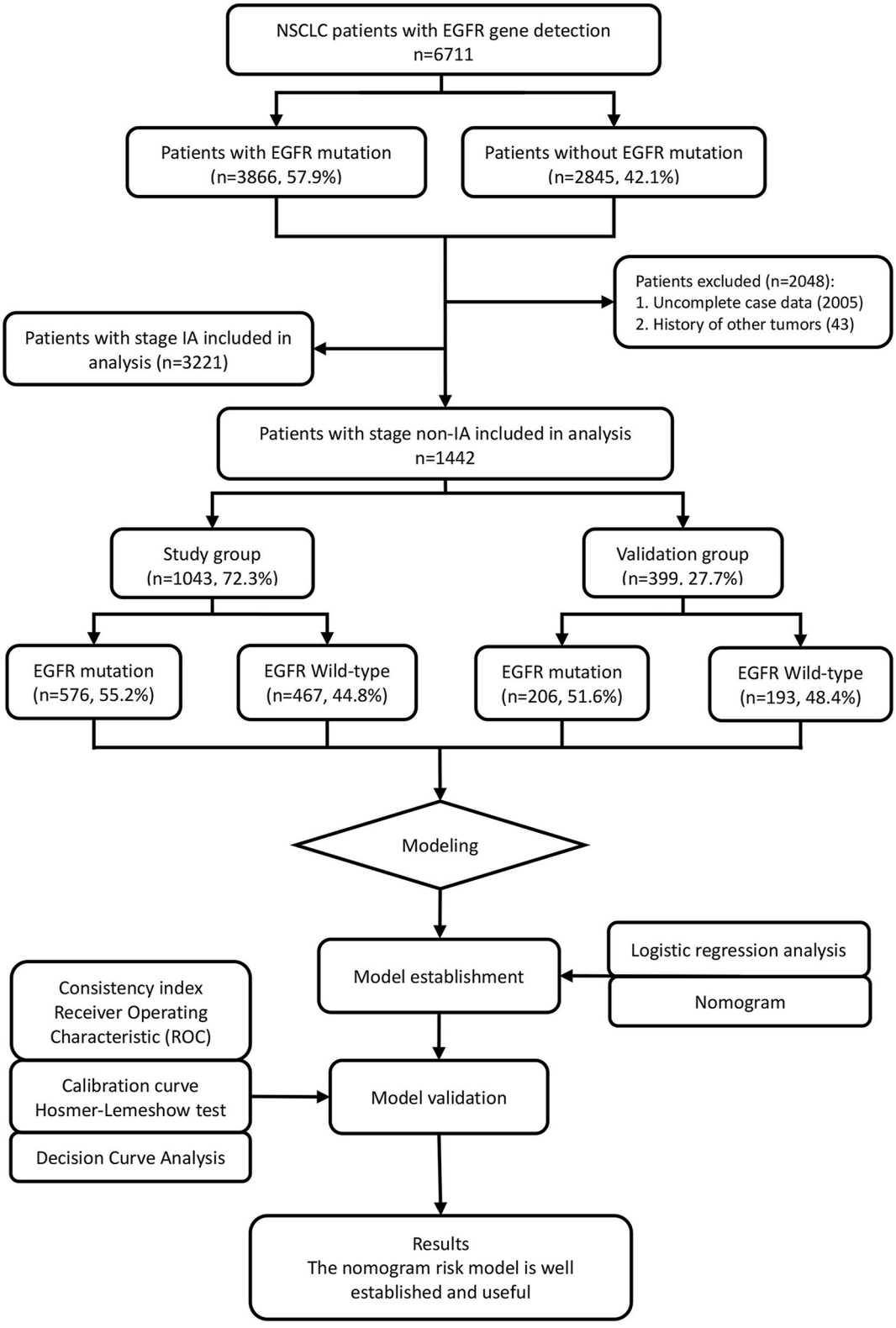

Fig. 1. Flow chart of the study design and analysis.

1. Introduction

Lung cancer is a prevalent malignant tumor and a leading cause of cancer-related mortality worldwide [1,2]. Non-small cell lung cancer (NSCLC) accounts for approximately 85 % of all lung cancer cases [3], with epidermal growth factor receptor (EGFR) mutations being the most common driver mutation in NSCLC [4]. These mutations occur in about 50 % of NSCLC patients in the Asia-Pacific region and 15 % of patients in Western countries [5,6]. Multiple clinical studies have unequivocally demonstrated that advanced NSCLC patients with EGFR mutations are more sensitive to treatment with EGFR tyrosine kinase inhibitors (EGFR-TKIs) as compared to traditional chemotherapy, especially those with advanced lung adenocarcinoma (ADC) [7–10]. EGFR-TKI monotherapy shows a higher overall response rate, longer median progression-free survival, and median overall survival. Furthermore, the incidence of treatment-related adverse events is significantly lower than that of chemotherapy [11].

However, due to the low detection rate of EGFR mutations in the real-world setting [12,13], many lung cancer patients are unable to benefit from this treatment approach, resulting in limited improvement in survival and quality of life. A systematic review of studies conducted worldwide to evaluate the utilization of EGFR mutation testing in routine care revealed that less than one-third of over 50, 000 patients from 18 eligible studies were tested for EGFR mutations [14]. Despite advances in genetic mutation testing methods, the main reasons for the lower-than-expected EGFR mutation detection rate are the lack of tumor tissue and the high cost of EGFR testing [12,15]. Therefore, there is an urgent need to develop a simple and non-invasive testing method to predict the EGFR mutation status and improve the detection rate of EGFR mutations.

Previous research has shown that serum tumor markers (STMs) can aid in the diagnosis of suspected clinical cancer and cancers of unknown primary origin. They may also play a significant role in cancer prognosis, treatment, and subsequent monitoring. The repertoire of currently employed biomarkers for primary lung cancer encompasses carcinoembryonic antigen (CEA), neuron-specific enolase (NSE), soluble fragment of cytokeratin 19 (CYFRA21-1), progastrin-releasing peptide (proGRP), squamous cell carcinoma antigen (SCC), and carbohydrate antigen 125 (CA125) [16–18]. Although the singular utility of individual tumor markers in terms of specificity and sensitivity remains somewhat limited, their combined application has emerged as a strategy to bolster diagnostic accuracy. Moreover, the association between lung cancer biomarkers and clinical staging is noteworthy, as lower levels or positive rates of SCC, CEA, NSE and CYFRA21-1 have been observed in patients diagnosed with early-stage NSCLC [19]. Previous studies have revealed the value of different serum markers in predicting the EGFR mutation status in NSCLC patients [20–25]. However, these research outcomes have exhibited inconsistencies and the impact of tumor staging has not been adequately addressed. The development of predictive models through traditional regression analysis or machine learning methods has enabled the integration of a multitude of parameters to provide individualized diagnostic predictions [26]. Limited reports exist regarding the prediction EGFR mutations in non-stage IA NSCLC patients using STMs. The model's performance in predicting EGFR mutations is deemed unsatisfactory. Therefore, we retrospectively characterized STMs and other clinical factors of non-stage IA NSCLC patients and explored their combined utility in developing nomogram and 8 machine learning models for accurately predicting EGFR mutations.

2. Materials and methods

2.1. Study design and patient cohort

We retrospectively collected a total of 6711 NSCLC patients who underwent EGFR gene testing at our institution between November 2016 and October 2020. Demographic information and clinical characteristics data were extracted and organized from electronic medical records of the patients. The inclusion criteria for this study were as follows: (1) patients must have undergone pretreatment testing for six STMs, including CEA, SCC, CYFRA 21-1, NSE, proGRP, and CA125, (2) complete information on demographic and clinical characteristics including age, gender, smoking history, and pathology. Patients with a history of other malignancies were excluded from the study. Subsequently, NSCLC patients were categorized into stage IA and non-stage IA based on tumor staging [27]. The non-stage IA patients were further divided into a study group (with gene testing dates ranging from November 2016 to December 2019) and a validation group (with gene testing dates from January 2020 to October 2020). Ultimately, the study group included 1043 patients, while the validation group consisted of 399 patients (Fig. 1). This study was conducted in accordance with the principles outlined in the Helsinki Declaration. The Institutional Review Board of the Affiliated Hospital of Qingdao University approved this retrospective study with the ethical approval number: QYFY WZLL 27853. Informed consent was waived by the institutional review board due to the retrospective nature of the study.

2.2. Histopathology examination

Hematoxylin and eosin (HE)-stained tumor slides derived from formalin-fixed paraffin-embedded tissues of tumor specimens were subjected to microscopic examination and assessment by two pathologists. Any discrepancies were resolved through consensus. Pathological histological subtype of the tumor was documented. The tumor stage was determined using the tumor-node-metastasis (TNM) staging system based on the 8th edition of the International Union against Cancer staging system [27]. The histological subtype was evaluated according to the 2015 World Health Organization (WHO) classification [28].

2.3. STMs measurement

Peripheral venous blood samples (3 mL) were collected from each patient for the detection of lung cancer-associated tumor

markers. Serum was separated by centrifugation at 3000×*g* for 10 min. The serum concentrations of tumor markers were measured using a commercial chemiluminescent immunoassay kit (MAGLUMI 4000 Plus, China). Blood samples from all participants were obtained via peripheral venous puncture prior to any anticancer treatment. The positivity threshold of STMs were as follows: CEA: 5 ng/mL, CA125: 35 U/mL, CYFRA21-1: 3.3 ng/mL, SCC: 2.5 ng/mL, NSE: 17 ng/mL, and ProGRP: 63 pg/mL. If the optimal cut-off (OCF) value based on the receiver operating characteristic (ROC) curve in the study exceeded the positivity threshold of the method, the OCF value was chosen as the final positivity threshold.

2.4. EGFR mutation analysis

Histological specimens of primary tumors, metastatic lymph nodes or organs, and cytological specimens of pleural or pericardial effusions were collected for EGFR gene testing. All samples were fixed in 10 % neutral buffered formalin and embedded in paraffin. Genomic DNA from tumor tissues or cells was extracted following the instructions of the Human EGFR Mutation Detection Kit (Amoy Diagnostics Co., Ltd., China). Polymerase chain reaction (PCR) was performed using the ABI 7500 fluorescence PCR system (Thermo Fisher Scientific, China). The amplification refractory mutation system (ARMS) of the Human EGFR Mutation Detection Kit was used to determine the EGFR mutation status. If any exon mutation was detected, the tumor was identified as "EGFR mutation"; otherwise, the tumor was identified as "EGFR wild-type".

2.5. Statistical analysis

Non-normally distributed continuous variables were represented using the median, and group comparisons were conducted using non-parametric tests. Categorical variables were expressed as proportions, and group comparisons were conducted using the chi-square test and Fisher's exact test. Factors that showed statistical significance in the univariate analysis were further analyzed using multiple logistic regression analysis. The effect measure of each variable on EGFR mutations was presented as odds ratios (OR) and corresponding 95 % confidence intervals (CI). Subsequently, the nomogram prediction model was developed utilizing the results of the multivariable analysis. The area under the curve (AUC) was calculated to assess the predictive performance of the model. The comparison of ROC curves was performed using the DeLong test. The clinical utility of the model was evaluated using Decision Curve Analysis [29]. Internal and external validation of the model was conducted through measures such as the concordance index (C-index), calibration curve, and Hosmer-Lemeshow test. Bootstrap resampling (1000 iterations) was employed to generate the calibration curve. In order to enhance the accuracy of predicting EGFR mutations, we employed 8 machine learning algorithms, including Random Forest (RF), Gradient Boosting Machine (GBM), Neural Network (NNET), Support Vector Machines (SVM), Lasso Regression algorithm (LASSO), Generalized Linear Model (GLM), K-Nearest Neighbor (KNN), and Logistic Regression (LR). For each model, C-index was computed separately for the training and test cohorts, and the model with the highest average C-index was deemed optimal. All p values were two-sided, and a p value less than 0.05 was considered to be statistically significant. Statistical analyses were performed with IBM SPSS Statistics version 25.0 (IBM Corp. New York, USA) and R (version 4.2.2, R Development Core Team), including the "pROC", "regplot", "rms" and "ResourceSelection" packages.

3. Results

3.1. Patient clinical characteristics

Among a total of 6,711 NSCLC patients who underwent EGFR gene testing, EGFR mutations were detected in 3,866 patients (57.9 %) (Fig. 1). Among patients with EGFR mutations, common mutations were observed in 3,370 cases (87.1 %), rare mutations in 330

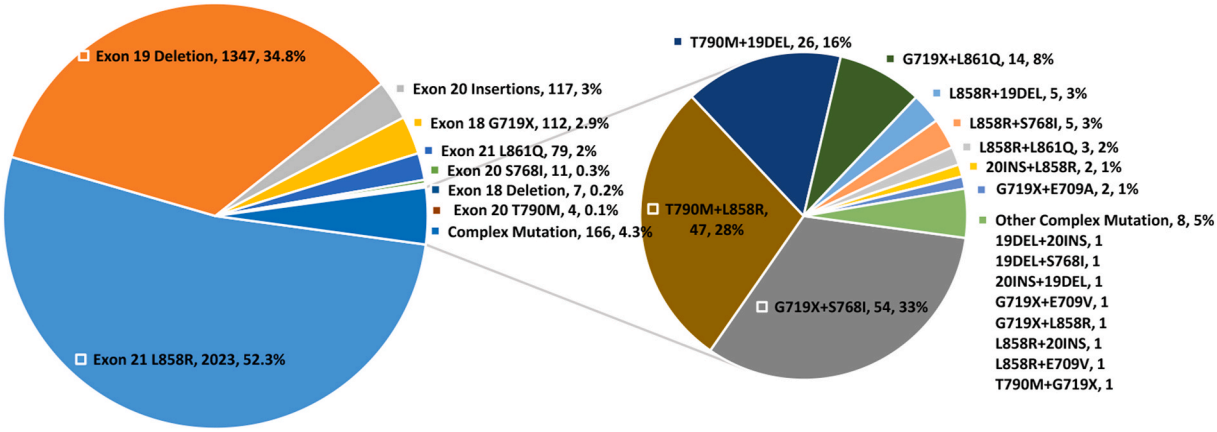

Fig. 2. Distribution of EGFR mutation subtypes in all NSCLC patients.

cases (8.6 %), and complex mutations in 166 cases (4.3 %). The common mutations included exon 19 deletions and exon 21 L858R mutations, accounting for 40 % and 30 % of EGFR mutations, respectively. The most frequent rare mutation was exon 20 insertion mutations. Among complex mutations, those with co-occurring exon 20 T790 M and exon 18 G719X mutations were the most prevalent, representing 44.6 % and 44 % of complex mutations. The specific numbers and frequencies of each EGFR mutation subtype are presented in Fig. 2.

A total of 3,221 NSCLC patients with stage IA were included in the analysis (Table 1). It is evident that patients with stage IA exhibited a relatively lower positivity rate of serum tumor markers, with CYFRA 21-1 having the highest positive rate at merely 20.9 %. EGFR mutations were more frequently detected in females (80 % vs. 57.6 %, $p < 0.001$), non-smokers (77.5 % vs. 53.6 %, $p < 0.001$), ADC (73.8 % vs. 10 %, $p < 0.001$), negative CEA (72.6 % vs. 66.9 %, $p = 0.021$), and negative SCC (72.5 % vs. 63.6 %, $p = 0.011$). In addition, out of 1442 patients with non-stage IA, 1043 patients were included in the study group. The OCF values for CEA and CYFRA 21-1 in our research exceeded the assay kit's positive thresholds. Therefore, for non-stage IA patients, CEA positivity was defined as CEA levels above 11.38 ng/mL, and CYFRA 21-1 positivity was defined as CYFRA 21-1 levels above 4.18 ng/mL when analyzing their relationship with EGFR mutations. Clinical characteristics of NSCLC patients, stratified by the EGFR mutation status in the study and validation cohorts, are summarized in Table 2. EGFR mutations were observed in 576 patients (55.2 %) in the study cohort and 206 patients (51.6 %) in the validation cohort. In the study cohort, EGFR mutations were more commonly found in females (77.7 % vs. 37.3 %, $p < 0.001$), non-smokers (69.7 % vs. 34.1 %, $p < 0.001$), ADC (62.2 % vs. 11.7 %, $p < 0.001$), positive CEA (65 % vs. 50.2 %, $p = 0.01$), negative CYFRA21-1 (62.5 % vs. 44.9 %, $p < 0.001$) and negative SCC (59.2 % vs. 28.4 %, $p < 0.001$). Furthermore, the validation group demonstrated clinical characteristics associated with EGFR mutations that were largely similar to those observed in the study group.

3.2. Exploration of risk factors for EGFR mutation

The results of both univariate and multivariate logistic regression analyses for predicting EGFR mutations in NSCLC patients with stage IA are presented in Table 3. In the multivariate analysis incorporating gender, smoking history, pathology, CEA, and SCC, female (OR, 2.001; $p < 0.001$), non-smoking (OR, 1.558; $p < 0.001$), and ADC (OR, 15.433; $p < 0.001$) were identified as independent risk factors for predicting EGFR mutations, while none of the STMs emerged as significant predictor. However, in the study cohort (Table 4), incorporating significant factors identified in the univariate analysis, including gender, smoking history, pathology, and STMs (CEA, SCC, and CYFRA 21-1), the multivariate analysis revealed that being female (OR, 3.318; $p < 0.001$), non-smoking (OR, 1.770; $p = 0.002$), ADC (OR, 6.767; $p < 0.001$), positive CEA (OR, 1.709; $p = 0.001$), and negative CYFRA 21-1 (OR, 0.541; $p < 0.001$)

Table 1
Clinical characteristics according to EGFR mutation in NSCLC patients with stage IA.

	All Patients (n = 3221)	EGFR Wild-type (n = 902)	EGFR Mutation (n = 2319)	P value
Gender				<0.001
Female	2069 (64.2)	413 (20)	1656 (80)	
Male	1152 (35.8)	489 (42.4)	663 (57.6)	
Age, year				0.159
Median (IQR)	60 (54–66)	60 (53–65)	60 (54–66)	
Smoking history				<0.001
Never	2476 (76.9)	556 (22.5)	1920 (77.5)	
Former/current	745 (23.1)	346 (46.4)	399 (53.6)	
CEA				0.021
Negative	2862 (88.9)	783 (27.4)	2079 (72.6)	
Positive	359 (11.1)	119 (33.1)	240 (66.9)	
CYFRA 21–1				0.201
Negative	2547 (79.1)	700 (27.5)	1847 (72.5)	
Positive	674 (20.9)	202 (30)	472 (70)	
SCC				0.011
Negative	3048 (94.6)	839 (27.5)	2209 (72.5)	
Positive	173 (5.4)	63 (36.4)	110 (63.6)	
NSE				0.398
Negative	2909 (90.3)	821 (28.2)	2088 (71.8)	
Positive	312 (9.7)	81 (26)	231 (74)	
ProGRP				0.484
Negative	3132 (97.2)	880 (28.1)	2252 (71.9)	
Positive	89 (2.8)	22 (24.7)	67 (75.3)	
CA125				0.354
Negative	3128 (97.1)	872 (27.9)	2256 (72.1)	
Positive	93 (2.9)	30 (32.3)	63 (67.7)	
Pathology				<0.001
Non-ADC	90 (2.8)	81 (90)	9 (10)	
ADC	3131 (97.2)	821 (26.2)	2310 (73.8)	

Values presented are n (%) and the positive thresholds for serum tumor markers are provided by the assay kits unless otherwise noted.
Abbreviations: IQR, interquartile range; CEA, Carcinoembryonic antigen; CYFRA21-1, Cytokeratin-19 fragment; SCC, Squamous cell carcinoma antigen; NSE, Neuron-specific enolase; proGRP, Progastrin-releasing peptide; CA125, Carbohydrate antigen 125; ADC, adenocarcinoma.

Table 2
Clinical characteristics according to EGFR mutation in NSCLC patients with non-stage IA.

Characteristics	Study cohort				Validation cohort			
	All Patients (n = 1043)	EGFR Wild-type (n = 467, 44.8 %)	EGFR Mutation (n = 576, 55.2 %)	P value	All Patients (n = 399)	EGFR Wild-type (n = 193, 48.4 %)	EGFR Mutation (n = 206, 51.6 %)	P value
Gender				<0.001				<0.001
Female	462 (44.3)	103 (22.3)	359 (77.7)		202 (50.6)	58 (28.7)	144 (71.3)	
Male	581 (55.7)	364 (62.7)	217 (37.3)		197 (49.4)	135 (68.5)	62 (31.5)	
Age, year				0.255				0.111
Median (IQR)	63 (55–68)	63 (56–68)	62 (55–67)		63 (55–68)	64 (57–68)	62 (53–68)	
Smoking history				<0.001				<0.001
Never	618 (59.3)	187 (30.3)	431 (69.7)		268 (67.2)	95 (35.4)	173 (64.6)	
Former/ current	425 (40.7)	280 (65.9)	145 (34.1)		131 (32.8)	98 (74.8)	33 (25.2)	
CEA [#]				<0.001				0.027
Negative	689 (66.1)	343 (49.8)	346 (50.2)		349 (87.5)	151 (51.7)	141 (48.3)	
Positive	354 (33.9)	124 (35)	230 (65)		292 (73.2)	42 (39.3)	65 (60.7)	
CYFRA 21–1 [a]				<0.001				0.002
Negative	613 (58.8)	230 (37.5)	383 (62.5)		278 (69.7)	120 (43.2)	158 (56.8)	
Positive	430 (41.2)	237 (55.1)	193 (44.9)		121 (30.3)	73 (60.3)	48 (39.7)	
SCC				<0.001				0.003
Negative	909 (87.2)	371 (40.8)	538 (59.2)		351 (88)	160 (45.6)	191 (54.4)	
Positive	134 (12.8)	96 (71.6)	38 (28.4)		48 (12)	33 (68.8)	15 (31.3)	
NSE				0.146				0.104
Negative	758 (72.7)	329 (43.4)	429 (56.6)		331 (83)	154 (46.5)	177 (53.5)	
Positive	285 (27.3)	138 (48.4)	147 (51.6)		68 (17)	39 (57.4)	29 (42.6)	
ProGRP				0.557				0.629
Negative	996 (95.5)	444 (44.6)	552 (55.4)		376 (94.2)	183 (48.7)	193 (51.3)	
Positive	47 (4.5)	23 (48.9)	24 (51.1)		23 (5.8)	10 (43.5)	13 (56.5)	
CA125				0.805				0.744
Negative	724 (69.4)	326 (45)	398 (55)		319 (79.9)	153 (48)	166 (52)	
Positive	319 (30.6)	141 (44.2)	178 (55.8)		80 (20.1)	40 (50)	40 (50)	
Pathology				<0.001				<0.001
Non-ADC	145 (13.9)	128 (88.3)	17 (11.7)		50 (12.5)	44 (88)	6 (12)	
ADC	898 (86.1)	339 (37.8)	559 (62.2)		349 (87.5)	149 (42.7)	200 (57.3)	

Values presented are n (%) and the positive thresholds for serum tumor markers are provided by the assay kits unless otherwise noted.
Abbreviations: IQR, interquartile range; CEA, Carcinoembryonic antigen; CYFRA21-1, Cytokeratin-19 fragment; SCC, Squamous cell carcinoma antigen; NSE, Neuron-specific enolase; proGRP, Progastrin-releasing peptide; CA125, Carbohydrate antigen 125; ADC, adenocarcinoma.
[a] The optimal cut-off values for CEA and CYFRA 21-1 were established at 11.38 ng/mL and 4.18 ng/mL, respectively, surpassing the positivity thresholds of the assay, and thus chosen as the ultimate positivity thresholds.

Table 3
Univariate and multivariate analyses of various predictive factors for EGFR mutation in NSCLC patients with stage IA.

Characteristics, Factor	Univariate analysis OR (95 % CI)	P value	Multivariate analysis[a] OR (95 % CI)	P value
Gender, Female	2.957 (2.523–3.467)	<0.001	2.001 (1.612–2.484)	<0.001
Age, Years	1.009 (1.001–1.018)	0.159		
Smoking history, Never	2.995 (2.521–3.557)	<0.001	1.558 (1.229–1.975)	<0.001
CEA, Positive	0.76 (0.601–0.96)	0.021	0.985 (0.765–1.268)	0.906
CYFRA 21–1, Positive	0.886 (0.735–1.067)	0.201		
SCC, Positive	0.663 (0.482–0.913)	0.011	0.965 (0.676–1.377)	0.844
NSE, Positive	1.121 (0.86–1.463)	0.398		
ProGRP, Positive	1.19 (0.731–1.938)	0.484		
CA125, Positive	0.812 (0.522–1.263)	0.354		
Pathology, ADC	25.323 (12.66–50.651)	<0.001	15.433 (7.639–31.178)	<0.001

The positive thresholds for serum tumor markers are provided by the assay kits unless otherwise noted.
Abbreviations: OR, odds ratio; 95 % CI, 95 % confidence interval; CEA, Carcinoembryonic antigen; CYFRA21-1, Cytokeratin-19 fragment; SCC, Squamous cell carcinoma antigen; NSE, Neuron-specific enolase; proGRP, Progastrin-releasing peptide; CA125, Carbohydrate antigen 125; ADC, adenocarcinoma.
[a] Items were included in the multivariate analysis only when the P value is < 0.05 in univariate analysis.

were independent risk factors for EGFR mutation, ultimately incorporated into the nomogram predictive model (Fig. 3A). Additionally, the predictive efficacy and clinical utility of these factors in predicting EGFR mutations were evaluated through ROC curve and decision curve analyses. The results revealed that the predictive model exhibited higher predictive efficacy for the EGFR mutations

Table 4
Univariate and multivariate analyses of various predictive factors for EGFR mutation in the study cohort.

Characteristics, Factor	Univariate analysis OR (95 % CI)	P value	Multivariate analysis[a] OR (95 % CI)	P value
Gender, Female	5.847 (4.436–7.706)	<0.001	3.318 (2.307–4.771)	<0.001
Age, Years	0.992 (0.979–1.006)	0.243		
Smoking history, Never	4.451 (3.418–5.795)	<0.001	1.770 (1.238–2.531)	0.002
CEA, Positive [b]	1.839 (1.411–2.396)	<0.001	1.709 (1.244–2.346)	0.001
CYFRA 21–1, Positive [b]	0.489 (0.381–0.628)	<0.001	0.541 (0.398–0.736)	<0.001
SCC, Positive	0.273 (0.183–0.407)	<0.001	0.792 (0.485–1.292)	0.35
NSE, Positive	0.817 (0.622–1.073)	0.146		
ProGRP, Positive	0.839 (0.467–1.507)	0.557		
CA125, Positive	1.034 (0.793–1.348)	0.805		
Pathology, ADC	12.416 (7.355–20.959)	<0.001	6.767 (3.812–12.013)	<0.001

The positive thresholds for serum tumor markers are provided by the assay kits unless otherwise noted.
Abbreviations: OR, odds ratio; 95 % CI, 95 % confidence interval; CEA, Carcinoembryonic antigen; CYFRA21-1, Cytokeratin-19 fragment; SCC, Squamous cell carcinoma antigen; NSE, Neuron-specific enolase; proGRP, Progastrin-releasing peptide; CA125, Carbohydrate antigen 125; ADC, adenocarcinoma.

[a] Items were included in the multivariate analysis only when the P value is < 0.05 in univariate analysis.

[b] The optimal cut-off values for CEA and CYFRA 21-1 were established at 11.38 ng/mL and 4.18 ng/mL, respectively, surpassing the positivity thresholds of the assay, and thus chosen as the ultimate positivity thresholds.

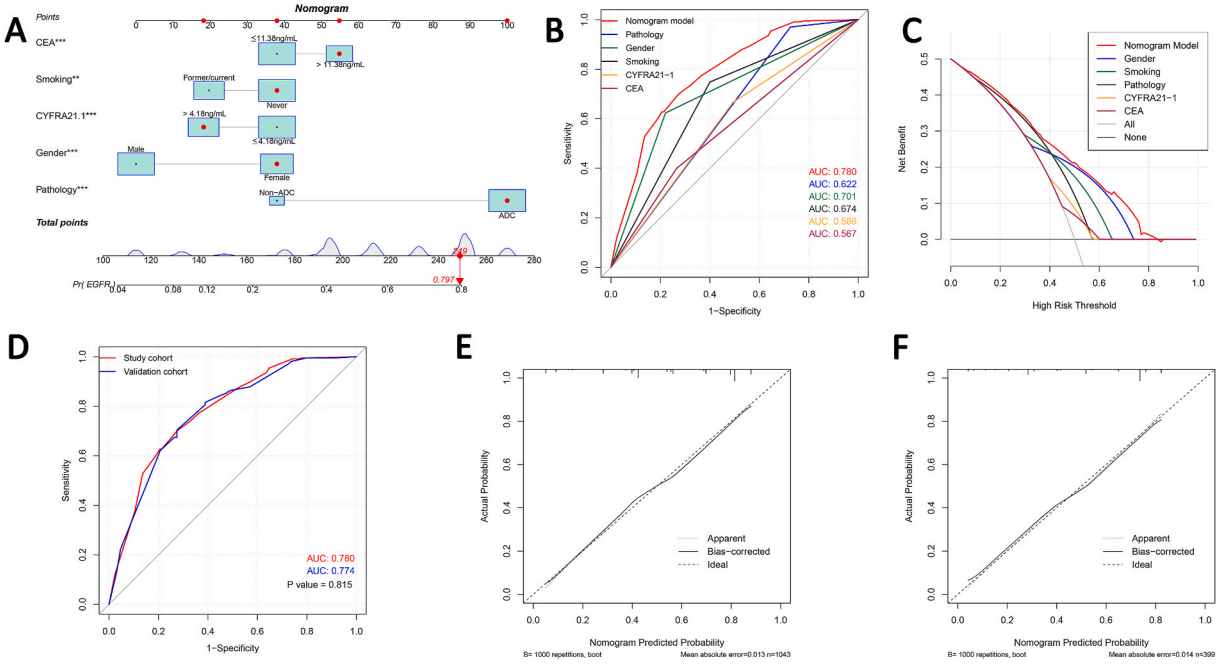

Fig. 3. Construction and validation of the nomogram predictive model. (A) The Nomogram model for predicting EGFR mutations in the study cohort. (B) ROC curves for the nomogram model in differentiating EGFR mutation status; (C) DCA curves to evaluate the clinical utility of the nomogram model for predicting EGFR mutations. (D) ROC curves for the discrimination of the nomogram; (E) The calibration plot in the study cohort; (F) The calibration plot in the validation cohort. Pr (EGFR): Probability of EGFR Mutation; ADC, adenocarcinoma; **means p < 0.01, ***means p < 0.001, ROC, receiver operating characteristic; DCA, decision curve analysis; AUC, area under the curve.

compared to individual factors, with AUC values of 0.780 (Fig. 3B). Moreover, decision curve analysis demonstrated that the net benefit of predicting EGFR mutations with the model surpassed that of individual factors (Fig. 4C). The probability threshold ranged of the model was 0–83 %, indicating a wider range and better clinical utility. Furthermore, for common EGFR mutations subtypes, CEA positivity and CYFRA 21-1 negativity were independent risk factors for both exon 19 deletion and exon 21 L858R mutations, nevertheless, negative SCC (OR, 0.446; p = 0.044) was an independent predictor of the exon 19 deletion mutation, while it did not predict the L858R mutation (Table 5).

3.3. Nomograms of the predictive model in the study cohort

Nomograms of the model was established based on the results of a multivariate analysis for predicting EGFR mutations (Fig. 3A).

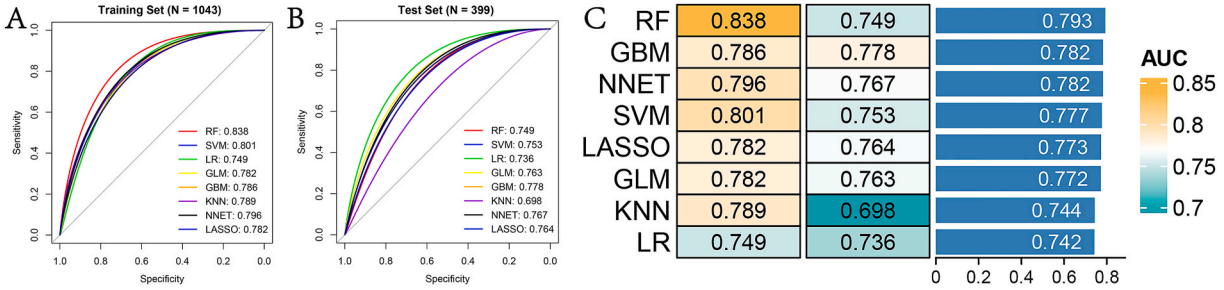

Fig. 4. ROC curves for 8 machine learning models in predicting EGFR mutations. (A) ROC curves in the study cohort; (B) ROC curves in the validation cohort; (C) A total of 8 kinds of prediction models and further calculated the C-index of each model. ROC, receiver operating characteristic; AUC, area under the curve; RF, Random Forest; GBM, Gradient Boosting Machine; NNET, Neural Network; SVM, Support Vector Machines; LASSO, Lasso Regression algorithm; GLM, Generalized Linear Model; KNN, K-Nearest Neighbor; LR, Logistic Regression.

Table 5
Multivariate analyses of various predictive factors for common EGFR mutation subtypes in the study cohort.

Characteristics, Factor	Exon 19 deletion mutation (n = 234) vs. EGFR Wild-type (n = 467)		Exon 21 L858R mutation (n = 273) vs. EGFR Wild-type (n = 467)	
	Multivariate analysis[a] OR (95 % CI)	P value	Multivariate analysis[a] OR (95 % CI)	P value
Gender, Female	3.021 (1.933–4.722)	<0.001	3.476 (2.264–5.335)	<0.001
Smoking history, Never	2.16 (1.351–3.453)	0.001	1.751 (1.128–2.719)	0.013
CEA, Positive [b]	1.799 (1.212–2.671)	0.004	1.687 (1.156–2.462)	0.007
CYFRA 21–1, Positive [b]	0.55 (0.372–0.813)	0.003	0.606 (0.418–0.878)	0.008
SCC, Positive	0.446 (0.203–0.978)	0.044	0.829 (0.452–1.522)	0.546
Pathology, ADC	6.066 (2.508–14.675)	<0.001	5.888 (2.784–12.455)	<0.001

The positive thresholds for serum tumor markers are provided by the assay kits unless otherwise noted.
Abbreviations: OR, odds ratio; 95 % CI, 95 % confidence interval; CEA, Carcinoembryonic antigen; CYFRA21-1, Cytokeratin-19 fragment; SCC, Squamous cell carcinoma antigen; NSE, Neuron-specific enolase; proGRP, Progastrin-releasing peptide; CA125, Carbohydrate antigen 125; ADC, adenocarcinoma.
[a] Items were included in the multivariate analysis only when the P value is < 0.05 in univariate analysis.
[b] The optimal cut-off values for CEA and CYFRA 21-1 were established at 11.38 ng/mL and 4.18 ng/mL, respectively, surpassing the positivity thresholds of the assay, and thus chosen as the ultimate positivity thresholds.

The probability of EGFR mutation can be assessed by assigning "Points" to each variable and summing them to obtain the total points. This total is then plotted on the "Total Points" axis, and a vertical line is drawn from the total points axis to the "Pr (EGFR)" axis. For instance, the probability of EGFR mutation was predicted in a female patient with ADC, positive CYFRA 21-1, positive CEA and non-smoking. The prediction scores were as follows: ADC scored 100, positive CYFRA 21-1 scored 18, positive CEA scored 55, non-smoking scored 38, and female scored 38. Upon summing these scores, the total reached 249, indicating the probability about 0.8 (80 %) for the presence of the EGFR mutation.

Table 6
Performances of discrimination and calibration of models in the study and validation cohorts.

Characteristics	Study cohort	Validation cohort
ROC analysis		
AUC/C-index	0.780	0.774
95 % CI	0.752–0.808	0.729–0.819
Sensitivity (%)	70	70.4
Specificity (%)	72.6	72.5
Calibration curves		
Corrected C-index	0.777	0.766
Mean absolute error	0.013	0.014
Hosmer-Lemeshow test		
X squared	0.826	2.599
P value[a]	0.662	0.273

Abbreviations: ROC, receiver operating characteristic; AUC, area under the curve; C-index, consistency index; 95 % CI, 95 % confidence interval.
[a] P value of the Hosmer-Lemeshow test >0.05 indicates that a model has high goodness of fit.

3.4. Performances of discrimination and calibration

The study and validation cohorts were utilized for the internal and external evaluation of model performance. The nomogram model achieved AUCs of 0.780 and 0.774 (P = 0.815) in the research and validation groups, respectively (Fig. 3D), indicating its favorable discrimination of EGFR mutations. The newly developed nomogram model was validated through internal (Fig. 3E) and external validation (Fig. 3F) using the bootstrap method with 1000-bootstrap repetitions, and the resulting calibration curves demonstrated strong consistency between the predicted values and the actual values. Furthermore, the calibrated C-index of 0.777 in the study cohort and 0.766 in the validation cohort, similar to the uncalibrated C-index (Table 6), indicated excellent predictive accuracy of the proposed nomogram model. Moreover, the Hosmer-Lemeshow goodness-of-fit test yielded non-significant results in both the study and validation groups, indicating no significant discrepancies between the predicted values and the actual values.

3.5. Prediction of EGFR mutations using 8 machine learning algorithms

The prediction of EGFR mutations in non-IA stage NSCLC patients was conducted using 8 machine learning algorithms, and C-index values were calculated for each model across the entire dataset. Results indicated that, irrespective of the training (Fig. 4A) or validation cohort (Fig. 4B), the RF model consistently exhibited the highest C-index values at 0.838 and 0.749, respectively. The RF model also demonstrated the highest average C-index of 0.793 among the eight machine learning algorithms, showcasing superior predictive efficiency (Fig. 4C). When compared to the nomogram model, the RF model exhibited even better predictive performance, with C-index values of 0.838 and 0.780 in the training cohort. Furthermore, the LR model, with the lowest average C-index among the machine learning models, still surpassed 0.74. This suggests that the 8 machine learning models, leveraging serum tumor markers and other clinical features, exhibit robust predictive efficacy for predicting EGFR mutations.

4. Discussion

EGFR-TKIs have demonstrated superior efficacy in the treatment of advanced lung cancer compared to chemotherapy, accompanied by a significant reduction in the incidence of treatment-related adverse reactions [4,30]. Biopsy-based EGFR gene testing currently serves as the gold standard for mutation detection. However, the low detection rate of EGFR mutations in real-world scenarios has limited the benefits of this treatment approach for many NSCLC patients [14]. Hence, there is an urgent need to develop a simple and noninvasive testing method to predict the EGFR mutation. Primary healthcare facilities can generally perform low-cost, rapid, and accurate detection of STMs. Therefore, we constructed novel predictive models consisting of STMs and other clinical features to predict the EGFR mutation in non-stage IA NSCLC patients using non-invasive, cost-effective, and readily accessible indicators. Additionally, our preliminary research phase collected thousands of EGFR gene testing results and provided detailed descriptions of the incidence of EGFR mutations and their subtypes, serving as a basis for future clinical investigations. To our knowledge, this is the first study to integrate easily obtainable clinical factors into the nomogram model and 8 machine learning algorithms for predicting the EGFR mutation in non-stage IA NSCLC patients.

Previous studies have revealed the value of different serum biomarkers in predicting the EGFR mutations in NSCLC patients. However, there are still some inconsistencies in the research findings. Arthur et al. [21] found no statistically significant differences in CEA, CYFRA21-1, or SCC levels between EGFR mutant and wild-type patients. Conversely, Jin et al. [20] reported an increase in the EGFR mutation rate with elevated CEA levels in non-smoking lung cancer patients. Wang et al. [24], in a study including 1089 patients, demonstrated an association between negative CYFRA21-1, negative SCC, negative CA125, and EGFR mutations in NSCLC patients. We thought that the discrepancies in these results could be attributed to the small sample sizes and the lack of consideration for the impact of tumor staging on STMs. Jiang et al. indicated an association between lung cancer biomarkers and tumor staging, with lower levels or positive rates of tumor markers observed in early-stage NSCLC patients [19]. Based on these considerations, our study collected a large amount of patient data and performed analyses based on whether the tumor staging was stage IA. The results revealed that the highest positive rate among the six STMs in stage IA NSCLC patients was only 20.9 %, most of which were below 10 %. In NSCLC patients with stage IA, multivariate analysis results showed that ADC, females, and non-smokers had a higher EGFR mutation rate, which is consistent with the non-stage IA patients in our study and the majority of research conclusions [31–33]. None of the six STMs was independent risk factor for predicting EGFR mutations, likely due to their low clinical value in the early stages of lung cancer. However, in patients beyond the stage IA, multivariate analysis results indicated that ADC, females, non-smokers, positive CEA and negative CYFRA21-1 were independent risk factors for predicting EGFR mutations. Subsequently, the predictive model was constructed using multivariate analysis results. ROC curves and decision curve analysis demonstrated that the model exhibited good predictive efficacy (AUC = 0.78) and clinical utility for predicting EGFR mutations. Furthermore, the predictive efficacy and clinical utility of the model were significantly superior to that of individual clinical features. Additionally, CEA positivity and CYFRA 21-1 negativity were independent risk factors for both exon 19 deletion and exon 21 L858R mutations. Considering that common EGFR mutations account for over 85 % of EGFR mutations in clinical practice, it is understandable that the independent risk factors for predicting the exon 19 deletion mutation and the exon 21 L858R mutation are essentially the same as those for predicting EGFR mutations. Nevertheless, negative SCC was an independent predictor of the exon 19 deletion mutation, while it did not predict the L858R mutation, which is similar to the findings of Wang et al. [24]. In a nutshell, STMs can predict the EGFR mutations in non-stage IA NSCLC patients, and the combination of STMs with other clinical factors can enhance the predictive efficacy and clinical utility for predicting EGFR mutations.

Therefore, we explored the development of the nomogram model combining STMs with other clinical features to provide

personalized risk assessment of EGFR mutations for non-IA NSCLC patients who were unable to undergo genetic testing. The nomogram model incorporating CEA, CYFRA21-1, pathology, gender, and smoking history for predicting EGFR mutations was constructed using the results of multivariate analysis. CYFRA 21-1 served as a tumor marker that exhibited enhanced sensitivity for NSCLC, particularly in squamous cell carcinoma [34]. CEA exhibits relatively high sensitivity in lung cancer, with the highest serum concentrations observed in ADC and large cell carcinoma [35]. EGFR mutations predominantly occur in patients with ADC, and approximately 40 % of lung ADC patients demonstrate elevated levels of CEA. Conversely, positive CYFRA 21-1 results are frequently associated with the presence of squamous cell carcinoma. This observation offers a potential explanation for the independent risk factors of positive CEA and negative CYFRA 21-1 in predicting EGFR mutations. In both internal and external validations, the calibration curves of the nomogram model clearly demonstrated a high degree of consistency between the predictions and observations. Furthermore, the calibrated C-index of 0.777 in the study cohort and 0.766 in the validation cohort, similar to the uncalibrated C-index, indicated excellent predictive accuracy of the proposed nomogram model. The results of the Hosmer-Lemeshow goodness-of-fit test were also non-significant ($P > 0.05$), indicating no significant differences between the predicted and actual values. Therefore, the use of nomogram is recommended. In other words, nomogram is almost accurate in predicting the probability of EGFR mutations in non-stage IA NSCLC patients. Furthermore, in our study, regardless of tumor stage, NSE, CA125, and proGRP showed no significant clinical significance in predicting EGFR mutations.

Random Forest, as a component of machine learning algorithms, has been applied in clinical outcome predictions [36]. RF constructs numerous decision trees through log-rank tests to identify different states and generates individual probabilities based on the average prediction results of all trees. Advantages of RF over traditional regression analysis include its unrestricted applicability and outstanding predictive performance. In our study, we tested eight machine learning models for predicting EGFR mutation occurrence, and RF demonstrated the optimal C-index, slightly surpassing the predictive efficacy of the nomogram model constructed through traditional regression analysis. Future endeavors will focus on leveraging artificial intelligence technologies to further enhance the clinical value of predictive models, facilitating non-invasive detection methods for predicting EGFR mutation occurrence and providing a basis for precision treatment for patients.

Our study also has several limitations. Firstly, being a retrospective study, there may exist patient inclusion and sample selection biases. Secondly, EGFR mutations are more prevalent in early-stage lung adenocarcinoma patients, prompting clinical physicians to prioritize genetic testing. Despite a slightly lower real-world occurrence rate than the finding of the study, the substantial sample size of this study reinforces the confidence in its accurate findings. Thirdly, the variation in EGFR mutation incidence across different regions may impact the clinical applicability of our results [5,6]. Fourthly, we did not assess treatment response or conduct survival analysis based on clinical features, including STM levels. Further large-scale prospective studies are warranted to validate our findings and explore the significance of monitoring treatment efficacy.

In conclusion, none of the STMs emerged as significant predictor for predicting EGFR mutations in stage IA NSCLC patients, while CYFRA21-1 and CEA have been identified as crucial factors for predicting EGFR mutations in non-stage IA patients. The nomogram and 8 machine learning models that combined STMs with other clinical factors could effectively predict the probability of EGFR mutations, providing valuable insights for personalized treatment.

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of the Affiliated Hospital of Qingdao University (QYFY WZLL 27853) and waived the need for informed consent due to the study design.

Funding

National Natural Science Foundation of China, Grant/Award Number: 82102188; Qingdao Postdoctoral Funding Project, Grant/Award Number: QDBSH20230101012.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

CRediT authorship contribution statement

Wenxing Du: Writing – review & editing, Writing – original draft, Visualization, Software, Formal analysis, Conceptualization. **Tong Qiu:** Writing – original draft, Methodology, Investigation, Conceptualization. **Hanqun Liu:** Writing – original draft, Methodology, Investigation, Conceptualization. **Ao Liu:** Writing – original draft, Methodology, Conceptualization. **Zhe Wu:** Writing – original draft, Investigation, Conceptualization. **Xiao Sun:** Writing – original draft, Methodology, Conceptualization. **Yi Qin:** Writing – original draft, Methodology, Conceptualization. **Wenhao Su:** Writing – original draft, Investigation, Conceptualization. **Zhangfeng Huang:** Writing – original draft, Investigation, Conceptualization. **Tianxiang Yun:** Writing – original draft, Investigation, Conceptualization. **Wenjie Jiao:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

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.

Abbreviations

STMs	Serum tumor markers
EGFR	Epidermal growth factor receptor
NSCLC	Non-small cell lung cancer
CEA	Carcinoembryonic antigen
SCC	Squamous cell carcinoma antigen
CYFRA21-1	Cytokeratin-19 fragment
AUC	Area under the curve
TKIs	Tyrosine kinase inhibitors
ADC	Adenocarcinoma
NSE	Neuron-specific enolase
proGRP	Progastrin-releasing peptide
CA125	Carbohydrate antigen 125
TNM	Tumor-node-metastasis
WHO	World Health Organization
OCF	Optimal cut-off
ROC	Receiver operating characteristic
PCR	Polymerase chain reaction
ARMS	Amplification refractory mutation system
OR	Odds ratios
CI	Confidence intervals
C-index	Concordance index
RF	Random Forest
GBM	Gradient Boosting Machine
NNET	Neural Network
SVM	Support Vector Machines
LASSO	Lasso Regression algorithm
GLM	Generalized Linear Model
KNN	K-Nearest Neighbor
LR	Logistic Regression

References

[1] R.L. Siegel, K.D. Miller, N.S. Wagle, A. Jemal, Cancer statistics, CA A Cancer J. Clin. 73 (1) (2023) 17–48, 2023.

[2] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (3) (2021) 209–249.

[3] R.S. Herbst, J.V. Heymach, S.M. Lippman, Lung cancer, N. Engl. J. Med. 359 (13) (2008) 1367–1380.

[4] J.A. Marin-Acevedo, B. Pellini, E.O. Kimbrough, J.K. Hicks, A. Chiappori, Treatment strategies for non-small cell lung cancer with common EGFR mutations: a review of the history of EGFR TKIs approval and emerging data, Cancers 15 (3) (2023).

[5] Y. Shi, J.S. Au, S. Thongprasert, S. Srinivasan, C.M. Tsai, M.T. Khoa, K. Heeroma, Y. Itoh, G. Cornelio, P.C. Yang, A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER), J. Thorac. Oncol. 9 (2) (2014) 154–162.

[6] A. Passaro, A. Prelaj, L. Bonanno, M. Tiseo, A. Tuzi, C. Proto, R. Chiari, D. Rocco, C. Genova, C. Sini, D. Cortinovis, S. Pilotto, L. Landi, C. Bennati, A. Camerini, L. Toschi, C. Putzu, G. Cerea, G. Spitaleri, F. Cappuzzo, F. de Marinis, Activity of EGFR TKIs in caucasian patients with NSCLC harboring potentially sensitive uncommon EGFR mutations, Clin. Lung Cancer 20 (2) (2019) e186–e194.

[7] Y.L. Wu, C.R. Xu, C.P. Hu, J. Feng, S. Lu, Y. Huang, W. Li, M. Hou, J.H. Shi, A. Marten, J. Fan, B. Peil, C. Zhou, Afatinib versus gemcitabine/cisplatin for first-line treatment of Chinese patients with advanced non-small-cell lung cancer harboring EGFR mutations: subgroup analysis of the LUX-Lung 6 trial, OncoTargets Ther. 11 (2018) 8575–8587.

[8] C. Zhou, Y.L. Wu, G. Chen, J. Feng, X.Q. Liu, C. Wang, S. Zhang, J. Wang, S. Zhou, S. Ren, S. Lu, L. Zhang, C. Hu, C. Hu, Y. Luo, L. Chen, M. Ye, J. Huang, X. Zhi, Y. Zhang, Q. Xiu, J. Ma, L. Zhang, C. You, Final overall survival results from a randomised, phase III study of erlotinib versus chemotherapy as first-line treatment of EGFR mutation-positive advanced non-small-cell lung cancer (OPTIMAL, CTONG-0802), Ann. Oncol. 26 (9) (2015) 1877–1883.

[9] J.C. Soria, Y. Ohe, J. Vansteenkiste, T. Reungwetwattana, B. Chewaskulyong, K.H. Lee, A. Dechaphunkul, F. Imamura, N. Nogami, T. Kurata, I. Okamoto, C. Zhou, B.C. Cho, Y. Cheng, E.K. Cho, P.J. Voon, D. Planchard, W.C. Su, J.E. Gray, S.M. Lee, R. Hodge, M. Marotti, Y. Rukazenkov, S.S. Ramalingam, F. Investigators, Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer, N. Engl. J. Med. 378 (2) (2018) 113–125.

[10] T.S. Mok, Y. Cheng, X. Zhou, K.H. Lee, K. Nakagawa, S. Niho, M. Lee, R. Linke, R. Rosell, J. Corral, M.R. Migliorino, A. Pluzanski, E.I. Sbar, T. Wang, J.L. White, Y.L. Wu, Improvement in overall survival in a randomized study that compared dacomitinib with gefitinib in patients with advanced non-small-cell lung cancer and EGFR-activating mutations, J. Clin. Oncol. 36 (22) (2018) 2244–2250.

[11] T.S. Mok, Y.L. Wu, M.J. Ahn, M.C. Garassino, H.R. Kim, S.S. Ramalingam, F.A. Shepherd, Y. He, H. Akamatsu, W.S. Theelen, C.K. Lee, M. Sebastian, A. Templeton, H. Mann, M. Marotti, S. Ghiorghiu, V.A. Papadimitrakopoulou, A. Investigators, Osimertinib or platinum-pemetrexed in EGFR t790m-positive lung cancer, N. Engl. J. Med. 376 (7) (2017) 629–640.
[12] Y. Cheng, Y. Wang, J. Zhao, Y. Liu, H. Gao, K. Ma, S. Zhang, H. Xin, J. Liu, C. Han, Z. Zhu, Y. Wang, J. Chen, F. Wen, J. Li, J. Zhang, Z. Zheng, Z. Dai, H. Piao, X. Li, Y. Li, M. Zhong, R. Ma, Y. Zhuang, Y. Xu, Z. Qu, H. Yang, C. Pan, F. Yang, D. Zhang, B. Li, Real-world EGFR testing in patients with stage IIIB/IV non-small-cell lung cancer in North China: a multicenter, non-interventional study, Thorac. Cancer 9 (11) (2018) 1461–1469.
[13] P.S. Aye, S. Tin Tin, M.J. McKeage, P. Khwaounjoo, A. Cavadino, J.M. Elwood, Development and validation of a predictive model for estimating EGFR mutation probabilities in patients with non-squamous non-small cell lung cancer in New Zealand, BMC Cancer 20 (1) (2020) 658.
[14] A.M. Thi, S. Tin Tin, M. McKeage, J.M. Elwood, Utilisation and determinants of epidermal growth factor receptor mutation testing in patients with non-small cell lung cancer in routine clinical practice: a global systematic review, Targeted Oncol. 15 (3) (2020) 279–299.
[15] Z. Lv, J. Fan, J. Xu, F. Wu, Q. Huang, M. Guo, T. Liao, S. Liu, X. Lan, S. Liao, W. Geng, Y. Jin, Value of (18)F-FDG PET/CT for predicting EGFR mutations and positive ALK expression in patients with non-small cell lung cancer: a retrospective analysis of 849 Chinese patients, Eur. J. Nucl. Med. Mol. Imag. 45 (5) (2018) 735–750.
[16] N. Vinolas, R. Molina, R. Fuentes, I. Bover, J. Rifa, V. Moreno, E. Canals, A. Marquez, E. Barreiro, J. Borras, X. Filella, J. Jo, X. Navarro, P. Viladiu, A.M. Ballesta, Tumor markers (CEA, CA 125, CYFRA 21.1, SCC and NSE) in non small cell lung cancer (NSCLC) patients as an aid in histological diagnosis and prognosis: comparison with the main clinical and pathological prognostic factors, Lung Cancer 29 (1, Supplement 1) (2000) 195.
[17] W. Qi, X. Li, J. Kang, Advances in the study of serum tumor markers of lung cancer, J. Cancer Res. Therapeut. 10 (Suppl) (2014) C95–C101.
[18] S. Cedres, I. Nunez, M. Longo, P. Martinez, E. Checa, D. Torrejon, E. Felip, Serum tumor markers CEA, CYFRA21-1, and CA-125 are associated with worse prognosis in advanced non-small-cell lung cancer (NSCLC), Clin. Lung Cancer 12 (3) (2011) 172–179.
[19] C. Jiang, M. Zhao, S. Hou, X. Hu, J. Huang, H. Wang, C. Ren, X. Pan, T. Zhang, S. Wu, S. Zhang, B. Sun, The indicative value of serum tumor markers for metastasis and stage of Non-small cell lung cancer, Cancers 14 (20) (2022).
[20] B. Jin, Y. Dong, H.M. Wang, J.S. Huang, B.H. Han, Correlation between serum CEA levels and EGFR mutations in Chinese nonsmokers with lung adenocarcinoma, Acta Pharmacol. Sin. 35 (3) (2014) 373–380.
[21] A. Cho, J. Hur, Y.W. Moon, S.R. Hong, Y.J. Suh, Y.J. Kim, D.J. Im, Y.J. Hong, H.J. Lee, Y.J. Kim, H.S. Shim, J.S. Lee, J.H. Kim, B.W. Choi, Correlation between EGFR gene mutation, cytologic tumor markers, 18F-FDG uptake in non-small cell lung cancer, BMC Cancer 16 (2016) 224.
[22] M. Jiang, P. Chen, X. Guo, X. Zhang, Q. Gao, J. Zhang, G. Zhao, J. Zheng, Identification of EGFR mutation status in male patients with non-small-cell lung cancer: role of (18)F-FDG PET/CT and serum tumor markers CYFRA21-1 and SCC-Ag, EJNMMI Res. 13 (1) (2023) 27.
[23] H. Zhang, M. He, R. Wan, L. Zhu, X. Chu, Establishment and evaluation of EGFR mutation prediction model based on tumor markers and CT features in NSCLC, J. Healthcare Eng. 2022 (2022) 8089750.
[24] S. Wang, P. Ma, G. Ma, Z. Lv, F. Wu, M. Guo, Y. Li, Q. Tan, S. Song, E. Zhou, W. Geng, Y. Duan, Y. Li, Y. Jin, Value of serum tumor markers for predicting EGFR mutations and positive ALK expression in 1089 Chinese non-small-cell lung cancer patients: a retrospective analysis, Eur. J. Cancer 124 (2020) 1–14.
[25] X. Tan, Y. Li, S. Wang, H. Xia, R. Meng, J. Xu, Y. Duan, Y. Li, G. Yang, Y. Ma, Y. Jin, Predicting EGFR mutation, ALK rearrangement, and uncommon EGFR mutation in NSCLC patients by driverless artificial intelligence: a cohort study, Respir. Res. 23 (1) (2022) 132.
[26] A. Rajkomar, J. Dean, I. Kohane, Machine learning in medicine, N. Engl. J. Med. 380 (14) (2019) 1347–1358.
[27] P. Goldstraw, K. Chansky, J. Crowley, R. Rami-Porta, H. Asamura, W.E. Eberhardt, A.G. Nicholson, P. Groome, A. Mitchell, V. Bolejack, S. International association for the study of lung cancer, A.B. Prognostic factors committee, I. Participating, S. International association for the study of lung cancer, B. Prognostic factors committee advisory, I. Participating, the IASLC lung cancer staging Project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer, J. Thorac. Oncol. 11 (1) (2016) 39–51.
[28] W.D. Travis, E. Brambilla, A.P. Burke, A. Marx, A.G. Nicholson, Introduction to the 2015 world Health organization classification of tumors of the lung, pleura, thymus, and heart, J. Thorac. Oncol. 10 (9) (2015) 1240–1242.
[29] A.J. Vickers, E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, Med. Decis. Making 26 (6) (2006) 565–574.
[30] A. Russo, T. Franchina, G. Ricciardi, A. Battaglia, M. Picciotto, V. Adamo, Heterogeneous responses to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) in patients with uncommon EGFR mutations: new insights and future perspectives in this complex clinical scenario, Int. J. Mol. Sci. 20 (6) (2019).
[31] S.P. D'Angelo, M.C. Pietanza, M.L. Johnson, G.J. Riely, V.A. Miller, C.S. Sima, M.F. Zakowski, V.W. Rusch, M. Ladanyi, M.G. Kris, Incidence of EGFR exon 19 deletions and L858R in tumor specimens from men and cigarette smokers with lung adenocarcinomas, J. Clin. Oncol. 29 (15) (2011) 2066–2070.
[32] R. Rosell, T. Moran, C. Queralt, R. Porta, F. Cardenal, C. Camps, M. Majem, G. Lopez-Vivanco, D. Isla, M. Provencio, A. Insa, B. Massuti, J.L. Gonzalez-Larriba, L. Paz-Ares, I. Bover, R. Garcia-Campelo, M.A. Moreno, S. Catot, C. Rolfo, N. Reguart, R. Palmero, J.M. Sanchez, R. Bastus, C. Mayo, J. Bertran-Alamillo, M. A. Molina, J.J. Sanchez, M. Taron, G. Spanish Lung Cancer, Screening for epidermal growth factor receptor mutations in lung cancer, N. Engl. J. Med. 361 (10) (2009) 958–967.
[33] M. Fukuoka, Y.L. Wu, S. Thongprasert, P. Sunpaweravong, S.S. Leong, V. Sriuranpong, T.Y. Chao, K. Nakagawa, D.T. Chu, N. Saijo, E.L. Duffield, Y. Rukazenkov, G. Speake, H. Jiang, A.A. Armour, K.F. To, J.C. Yang, T.S. Mok, Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS), J. Clin. Oncol. 29 (21) (2011) 2866–2874.
[34] A. Jafari-Kashi, H.A. Rafiee-Pour, M. Shabani-Nooshabadi, A new strategy to design label-free electrochemical biosensor for ultrasensitive diagnosis of CYFRA 21-1 as a biomarker for detection of non-small cell lung cancer, Chemosphere 301 (2022) 134636.
[35] A. Iwasaki, T. Shirakusa, Y. Yoshinaga, S. Enatsu, M. Yamamoto, Evaluation of the treatment of non-small cell lung cancer with brain metastasis and the role of risk score as a survival predictor, Eur. J. Cardio. Thorac. Surg. 26 (3) (2004) 488–493.
[36] D. Tian, H.J. Yan, H. Huang, Y.J. Zuo, M.Z. Liu, J. Zhao, B. Wu, L.Z. Shi, J.Y. Chen, Machine learning-based prognostic model for patients after lung transplantation, JAMA Netw. Open 6 (5) (2023) e2312022.