Nomogram Based on Preoperative CT Imaging Predicts The EGFR Mutation Status in Lung Adenocarcinoma

Guojin Zhang
Lanzhou University
Jing zhang
Lanzhou University
Yuntai Cao
Lanzhou University
Zhiyong Zhao
Lanzhou University
Shenglin Li
Lanzhou University
Liangna Deng
Lanzhou University
Junlin Zhou (ery_zhoujl@lzu.edu.cn)
Lanzhou University Second Hospital https://orcid.org/0000-0001-8109-6347

Research

Keywords: Adenocarcinoma, computed tomography, EGFR mutation, lung cancer

Posted Date: September 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-80072/v1

Version of Record: A version of this preprint was published at Translational Oncology on January 1st, 2021. See the published version at https://doi.org/10.1016/j.tranon.2020.100954.

Abstract

Background: Tyrosine kinase inhibitors (TKIs) provide clinical benefits to the lung cancer patients with epidermal growth factor receptor (EGFR) mutations. However, non-invasively determine EGFR mutation status in patients before targeted therapy remains a challenge. This study aimed to develop and validate a nomogram for preoperative prediction of EGFR mutation status in patients with lung adenocarcinoma.

Methods: This study retrospectively collected medical records of 403 patients with histologically confirmed lung adenocarcinoma from January 2016 and June 2020. The patients were divided into development and validation cohorts. The preoperative information on all patients was obtained, including clinical characteristics and computed tomography (CT) features. Multivariate logistic regression analysis was used to develop the predictive model. We combined CT features and clinical risk factors and used them to build a prediction nomogram. The performance of the nomogram was evaluated in terms of calibration, discrimination, and clinical usefulness. The nomogram was further validated in an independent external cohort.

Results: The predictive factors incorporated in the personalized prediction nomogram included smoking history (OR, 0.2; 95% CI: 0.1, 0.4; $P < 0.001$), bubble-like lucency (OR, 2.2; 95% CI: 1.3, 3.8; $P = 0.003$), pleural attachment (OR, 0.4; 95% CI: 0.2, 0.7, $P = 0.001$) and thickened adjacent bronchovascular bundles (OR, 3.1; 95% CI: 1.8, 5.3; $P < 0.001$). Based on these parameters, the prediction model has good discrimination and calibration ability. The area under the curve in the development and validation cohorts were 0.784 (95% CI: 0.733, 0.835) and 0.740 (95% CI: 0.643, 0.838), respectively. Decision curve analysis showed that the model was clinically useful.

Conclusions: This study presented a nomogram that contained CT features and clinical risk factors, which could conveniently and non-invasively predict EGFR mutation status in patients with lung adenocarcinoma before surgery.

Background

In the past decade, with the discovery of a series of oncogenic driver genes in lung cancer, especially in patients with epidermal growth factor receptor (EGFR) mutations, tyrosine kinase inhibitors (TKIs), such as gefitinib and osimertinib, can provide significant clinical benefits; compared with platinum-based chemotherapy alone, approximately 70% of patients treated with TKIs achieved objective remission, improved progression-free survival, and improved quality of life[1–4]. Therefore, it is particularly important to determine the EGFR mutation status of patients with lung cancer before initiating TKIs therapy. Currently, the detection of EGFR status in tissue samples is the most commonly used and accurate method, however, due to small samples and sampling errors, it is not always possible to determine the EGFR mutation status before targeted therapy[5, 6]. In addition, biopsy increases the risk of local tumor metastasis, and repeated sampling increases the financial burden [7, 8]. Given these limitations, a noninvasive, easy to operate, and economical and practical technique for detecting EGFR mutation status is urgently needed.

Computed tomography (CT) is currently the most commonly used and most important imaging method in lung cancer diagnosis, clinical staging, efficacy evaluation, and follow-up after treatment [9, 10]. Some studies have used CT imaging features to predict EGFR mutation status and obtained good results [5, 11–17]. Suh et al. studied the relationship between CT features and EGFR mutation status and found that tumors with EGFR mutations had higher pure ground-glass opacity (GGO) and part-solid nodule (PSN) and smaller tumor size and higher proportion of inner solid portion, compared to those with wild-type EGFR [11]. Liu et al. analyzed the CT features of tumors in 385 patients with lung adenocarcinoma, and the results showed that the presence of EGFR mutations was significantly correlated with 14 CT features, such as smaller tumor, GGO, bubble-like lucency, pleural retraction, etc. Moreover, when conducting ROC analysis, CT features combined with clinical variables were found to have good diagnostic performance in predicting EGFR mutation status [5]. Although, these studies have found a correlation between CT features and EGFR mutation status, few studies have independently verified these results to confirm their reliability and accuracy.

The aim of this study was to develop and validate a nomogram that combined the CT features and clinical variables for predicting of EGFR mutation status in patients with lung adenocarcinoma.

Methods

Patient selection

This retrospective study was approved by the Institutional Review Committee of Lanzhou University Second Hospital (Lanzhou, China), and the informed patient consent was waived. Patients who met the following inclusion criteria were enrolled in this study. (1) Histologically confirmed as lung adenocarcinoma; (2) Detection of EGFR mutation status on tumor specimens; (3) Obtained pre-operative CT scan date; (4) No previous history of malignant tumor. Patients were excluded if (1) There was a lack of clinical data, such as age, sex, and smoking history; (2) Received treatment before surgery; (3) Younger than 18 years of age; (4) The interval between CT scan and surgery exceeded 1 month; (5) The CT image quality was poor. Finally, 403 patients were included in this study (mean age ± SD, 57.4 ± 9.45 years; 205 males, 198 females). The sample included 208 mutated and 195 wild-type EGFR cases. 302 patients (159 mutated and 143 wild-type EGFR cases) diagnosed with lung adenocarcinoma from January 2016 to March 2019 were included in the development cohort, and 101 patients (49 mutated and 52 wild-type EGFR cases) diagnosed with lung adenocarcinoma from April 2019 to June 2020 were included in the validation cohort. Supplementary Table S1 summarizes patient characteristics in the development and validation cohorts.

CT image acquisition

CT scans were performed using Discovery CT750 HD (GE Healthcare), Philips iCT 256 (Koninklijke Philips N.V.) and Somatom Sensation 64 (Siemens, Erlangen, Germany). Scanning parameters were as follows: tube voltage, 120 kVp; tube current, 150-200 mA; the layer thickness of the axial image was 5-mm and the

layer spacing was 5-mm; the reconstruction layer thickness was 1.25-mm and the reconstruction interval was 1.25-mm.

CT image interpretation

A radiologist with 5 years' experience (G.J.Z) in CT imaging of thoracic malignancies and another radiologist with 15 years' experience (J.Z) independently evaluated CT images. The two radiologists were blinded to clinical data and histological results. If there was any disagreement, a unanimous conclusion was reached after discussion. All CT images were read with both lung (width: 1500 to 2000 HU; level: -450 to -600 HU) and mediastinal (width: 300 to 350 HU; level: 30 to 50 HU) window settings. In terms of morphological characteristics, a total of 20 CT features were assessed (Table 1). Spiculation was defined as a linear image extending from the edge of a nodule or mass into the lung parenchyma without being associated with the pleura [18]. Air bronchogram was defined as the pattern of air bronchi under an opaque background [19]. Bubblelike lucency was defined as a small spot of air attenuation with a diameter of less than 5 mm in a mass or nodule [16].

EGFR mutationanalysis

EGFR mutations were detected in tumors that were histologically confirmed as lung adenocarcinoma. Analysis of mutation status of EGFR exons 18~21 was examined using a PCR–based amplified refractory mutation system (ARMS) by using the human EGFR gene detection kit (Beijing SinoMD Gene Detection Technology Co., Ltd, Beijing, China). If exons 18, 19, 20, and 21 were mutated at any point, EGFR mutation was considered; otherwise, EGFR was defined as wild-type.

Statistical analyses

All statistical analyses were performed using R software (version 3.6.0). Countable variables were expressed as percentages and measurement data as mean ± standard deviation (SD). Univariate and multivariate logistic regression were used to analyze clinical and CT variables. Multivariate logistic regression analysis was performed for variables with differences in univariate analysis, and the stepwise forward method was used to select the variables ultimately included in the model. Two-sided $P < 0.05$ indicated significant difference.

Based on the variables of multiple logistic regression, we established an individualized nomogram prediction model related to EGFR mutation status. We drew a calibration curve to evaluate the goodness of fit of the prediction model. The ability of the prediction model was evaluated according to the discrimination and calibration. The distinction of the predictive model refers to its ability to distinguish patients with mutated EGFR from patients with wild-type EGFR. Decision curve analysis was performed to determine the clinical benefit of the nomogram. Use the area under the curve (AUC) of the receiver operating characteristic (ROC) curve to evaluate the discrimination of the dichotomy result [20]. AUC values ranged from 0.5 to 1. The larger the AUC value was, the better was the discrimination ability of the prediction model. Generally, AUC > 0.75 was considered to have excellent discrimination [21].

Results

Population characteristics

A total of 403 patients were enrolled, including 302 in the development cohort (mean age ± SD, 57.97 ± 9.53 years, 148 males and 154 females) and 101 in the validation cohort (mean age ± SD, 56.54 ± 9.20 years, 53 males and 48 females). The EGFR mutation rate was 52.6% and 48.5% in the development and validation cohorts (P = 0.472), respectively. There were no significant differences between the two cohorts in age (P = 0.191), sex (P = 0.546), smoking history (P = 0.35), carcinoembryonic antigen (CEA) (P=0.433), EGFR mutation status (P=0.472), and CT features (including 20 CT features) (P > 0.05)(Supplementary table S1). The balance between the two cohorts illustrated the rationality in patient grouping in this study.

The relationship between the characteristics of patients in the development cohort (including clinical characteristics and CT features) and EGFR mutation status are shown in Table 1. There was no statistical difference in age (57.55 ± 9.31 vs. 58.43 ± 9.78; odds ratio [OR], 0.99; 95% confidence interval [CI]: 0.97, 1.01; P = 0.421) and CEA (72.3% vs. 66.4%; OR, 1.32; 95% CI: 0.81, 2.16; P = 0.267) between patients with mutated EGFR and wild-type EGFR. Patients with EGFR mutations were more likely to be females (63.5% vs. 37.1%; OR, 2.296; 95% CI: 1.85, 4.72; P < 0.001) and have never smoking status (85.5% vs. 59.4%; OR, 0.25; 95% CI: 0.14, 0.43; P < 0.001) compared with those with wild-type EGFR.

Prediction nomogram construction

Univariate analysis of the development cohort showed that clinical factors with statistically significant differences in EGFR mutation status included sex and smoking history, and CT features factors included long-axis diameter (OR, 0.77; 95% CI: 0.69, 0.86; P < 0.001), size (OR, 0.46; 95% CI: 0.28, 0.77, P = 0.003), contour (OR, 0.48; 95% CI: 0.30, 0.79, P = 0.003), speculation (OR, 2.42; 95% CI: 1.49, 3.95, P <0.001), air bronchogram (OR, 2.03; 95% CI: 1.28, 3.22, P = 0.003), bubble-like lucency (OR, 2.78; 95% CI: 1.74, 4.43, P < 0.001), calcification (OR, 0.53; 95% CI: 0.30, 0.94, P = 0.029), vascular convergence (OR, 2.89; 95% CI: 1.60, 5.22, P < 0.001), pleural attachment (OR, 0.36; 95% CI: 0.22, 0.59, P < 0.001), thickened adjacent bronchovascular bundles (OR, 4.19; 95% CI: 2.58, 6.79, P < 0.001), and pleural effusion (OR, 0.59; 95% CI: 0.36, 0.96, P = 0.035), whereas age, CEA, distribution, lobe location, short-axis diameter, lobulation, texture, fissure attachment, peripheral emphysema, peripheral fibrosis, and inflammation were not related to EGFR mutation status (P > 0.05) (Table 1).

Multivariate logistic regression analysis was performed on variables with statistical differences in the univariate analysis and found that smoking history (OR, 0.2; 95% CI: 0.1, 0.4; P < 0.001), bubble-like lucency (OR, 2.2; 95% CI: 1.3, 3.8, P = 0.003), pleural attachment (OR, 0.4; 95% CI: 0.2, 0.7, P = 0.001), and thickened adjacent bronchovascular bundles (OR, 3.1; 95% CI: 1.8, 5.3; P < 0.001) were independent risk factors related to EGFR mutation status (Table 1).

Based on the results of multiple logistic regression analysis, we used four independent risk factors to generate an individualized nomogram to predict the EGFR mutation status (Figure 1). In the development

(OR, 1.08; 95% CI: 0.86, 1.36) and the validation (OR, 0.94; 95% CI: 0.61, 1.46) cohorts, the 95% CI of the calibration curve did not cross the diagonal line (Figures 2a and 2b). Therefore, the predicted probability of the nomogram model was consistent with the actual probability, which indicated that the model had good consistency.

Performance comparison of four risk factors and nomograms

The ROC curves of the four independent risk factors and the hybrid nomogram are shown in Figure 3. In the development cohort, the AUC of the hybrid nomogram was 0.784 (95% CI: 0.733, 0.835), while the AUC of smoking history, bubble-like lucency, pleural attachment, and thickened adjacent bronchovascular bundles were 0.630 (95% CI: 0.582, 0.679), 0.625 (95% CI: 0.570, 0.679), 0.616 (95% CI: 0.563, 0.760) and 0.671 (95% CI: 0.618, 0.724), respectively. (Figure 3a, Table 2). The AUC of the external validation cohort was 0.740 (95% CI: 0.643, 0.838) (Figure 3b, Table 2).

Decision curve analysis

The decision curves of the four risk factors and the hybrid nomogram are shown in Figure 4. Compared with the four risk factors, when the threshold probability was less than 83%, the hybrid nomogram showed the highest net benefit.

Discussion

In this study, we developed and validated a nomogram based on CT features including bubble-like lucency, pleural attachment, and thickened adjacent bronchovascular bundles and clinical variable, the smoking history, for personalized prediction of EGFR mutation status. The nomogram development cohort included 302 patients with lung adenocarcinoma. To further validate the performance of this model, we evaluated it in an independent external validation cohort that included 101 cases. The AUC of the model in the development and validation cohorts was 0.784 and 0.740, respectively.

EGFR mutation rate in this study was 51.6% (208/403), consistent with that shown in previous reports [5, 12, 15, 16, 22]. The EGFR mutation rate in the development cohort was slightly higher than that in the validation cohort (52.6% [159/302] vs. 48.5% [49/101]), and there was no statistical difference between the two cohorts (P = 0.472). The reason for this fluctuation may be related to our data selection bias. In addition, we found that female sex and non-smoking status were more common in the group with mutated EGFR than in the wild-type group, which was in agreement with previously published reports [5, 11, 15–17, 23–25]. Several studies have identified CT features associated with EGFR mutation status [5, 11–17, 26–29]. Consistent with the previous results, we found in multivariate analysis that smoking history, bubble-like attachment, and thickened bronchovascular bundles were independently correlated with EGFR mutations [5, 11, 16].

For the construction of the nomogram, we carried out multiple logistic regression analyses on the variables that showed statistical differences in univariate analysis. We then selected the variables that

showed significant differences in the univariate analysis in the final model. Finally, we used 13 candidate variables, including 11 CT features and 2 clinical variables, which were reduced to 4 potential predictor variables including 3 CT features and 1 clinical variable. This method not only simplified the method of selecting a predictor variable based on the strength of the predictor variable's univariate association with the result [30], but also presented the result variable related to the predictor variable to the observer more intuitively. Girard et al. used the clinical and pathological data of non-small cell lung cancer patients to establish a nomogram model to predict the EGFR mutation status, and the AUC of the model in the independent validation dataset reached 0.84 [31]. However, since the model included histological subtypes of tumors, this variable was obtained after that surgery and was invasive. Therefore, that model by Girard et al. had certain limitations in predicting the EGFR mutation status before surgery. In the present study, the variables included in the model were easily available before surgery and were non-invasive. Therefore, our model is easier to use in clinical practice.

One of the most important reasons for using a nomogram is that it can explain the needs of individual treatment or care needs. However, the clinical consequences of a specific level of discrimination or misalignment cannot be captured by performance, discrimination, and calibration of risk prediction [32–34]. Therefore, in order to demonstrate the usefulness of a nomogram, a decision curve was used to evaluate whether the decision of a nomogram model would improve the prognosis of patients. This novel approach analyzed clinical consequences based on threshold probabilities and derived net benefits from it [35–37]. The decision curve showed that when the threshold probability was < 83%, using the model to predict EGFR mutation status brought greater net benefit to TKIs therapy.

Our nomogram model still has some limitations. Firstly, this was a single-center retrospective study. Although we conducted independent external validation of the model, multi-center validation is still needed in subsequent studies to confirm the practicality of this model. Secondly, our analysis was limited to lung adenocarcinoma and did not involve other histological subtypes, because most EGFR mutations are found in lung adenocarcinoma.

Conclusion

Our study constructed a personalized nomogram model incorporating CT features and clinical risk factors, which can conveniently and non-invasively predict EGFR mutation status before surgery.

Abbreviations

ARMA, amplified refractory mutation system; AUC, area under the curve;

CEA, carcinoembryonic antigen; CT, computed tomography;

GGO, ground-glass opacity; EGFR, epidermal growth factor receptor;

ROC, receiver operating characteristic curve; TKIs, tyrosine kinase inhibitors.

Declarations

Ethics approval and consent to participate:

This study was approved by the Institutional Review Committee of Lanzhou University Second Hospital (Lanzhou, China), and the informed patient consent was waived.

Consent for publication:

Not applicable.

Availability of data and materials:

All the data presented is available upon request.

Competing interests:

The authors declare that they have no conflicts of interest.

Funding: This work was supported by the Talent Innovation and Entrepreneurship Project of Lanzhou (grant number 2016-RC-58); Open Fund project of Key Laboratory of Medical Imaging of Gansu Province (grant number GSYX202010).

Authors' contributions

All authors had access to the study data. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Sequist LV, Han J-Y, Ahn M-J, et al. Osimertinib plus savolitinib in patients with EGFR mutation-positive, MET-amplified, non-small-cell lung cancer after progression on EGFR tyrosine kinase inhibitors: interim results from a multicentre, open-label, phase 1b study. *Lancet Oncol* 2020, 21:373-386.
2. Robichaux JP, Elamin YY, Tan Z, et al. Mechanisms and clinical activity of an EGFR and HER2 exon 20-selective kinase inhibitor in non-small cell lung cancer. *Nat Med* 2018, 24:638-646.
3. Wu S-G, Shih J-Y. Management of acquired resistance to EGFR TKI-targeted therapy in advanced non-small cell lung cancer. *Mol Cancer* 2018, 17:38.
4. Rosell R, Carcereny E, Gervais R, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* 2012, 13:239-246.

5. Liu Y, Kim J, Qu F, et al. CT Features Associated with Epidermal Growth Factor Receptor Mutation Status in Patients with Lung Adenocarcinoma. *Radiology* 2016, 280:271-280.
6. Sacher AG, Dahlberg SE, Heng J, et al. Association Between Younger Age and Targetable Genomic Alterations and Prognosis in Non-Small-Cell Lung Cancer. *JAMA Oncol* 2016, 2:313-320.
7. Loughran CF, Keeling CR. Seeding of tumour cells following breast biopsy: a literature review. *Br J Radiol* 2011, 84:869-874.
8. Rios Velazquez E, Parmar C, Liu Y, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res* 2017, 77:3922-3930.
9. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017, 14:749-762.
10. Kauczor H-U, Heussel CP, von Stackelberg O. Time to take CT screening to the next level? *Eur Respir J* 2017, 49.
11. Suh YJ, Lee H-J, Kim YJ, et al. Computed tomography characteristics of lung adenocarcinomas with epidermal growth factor receptor mutation: A propensity score matching study. *Lung Cancer* 2018, 123:52-59.
12. Hasegawa M, Sakai F, Ishikawa R, et al. CT Features of Epidermal Growth Factor Receptor-Mutated Adenocarcinoma of the Lung: Comparison with Nonmutated Adenocarcinoma. *J Thorac Oncol* 2016, 11:819-826.
13. Jeon JH, Kang CH, Kim H-s, et al. Prognostic and predictive role of epidermal growth factor receptor mutation in recurrent pulmonary adenocarcinoma after curative resection. *Eur J Cardiothorac Surg* 2015, 47:556-562.
14. Usuda K, Sagawa M, Motono N, et al. Relationships between EGFR mutation status of lung cancer and preoperative factors - are they predictive? *Asian Pac J Cancer Prev* 2014, 15:657-662.
15. Lee H-J, Kim YT, Kang CH, et al. Epidermal growth factor receptor mutation in lung adenocarcinomas: relationship with CT characteristics and histologic subtypes. *Radiology* 2013, 268:254-264.
16. Choi C-M, Kim MY, Hwang HJ, et al. Advanced adenocarcinoma of the lung: comparison of CT characteristics of patients with anaplastic lymphoma kinase gene rearrangement and those with epidermal growth factor receptor mutation. *Radiology* 2015, 275:272-279.
17. Yano M, Sasaki H, Kobayashi Y, et al. Epidermal growth factor receptor gene mutation and computed tomographic findings in peripheral pulmonary adenocarcinoma. *J Thorac Oncol* 2006, 1:413-416.
18. Zwirewich CV, Vedal S, Miller RR, et al. Solitary pulmonary nodule: high-resolution CT and radiologic-pathologic correlation. *Radiology* 1991, 179:469-476.
19. Marten K, Engelke C. Computer-aided detection and automated CT volumetry of pulmonary nodules. *Eur Radiol* 2007, 17:888-901.
20. Harrell FE, Califf RM, Pryor DB, et al. Evaluating the yield of medical tests. *JAMA* 1982, 247:2543-2546.
21. Niu XK, He WF, Zhang Y, et al. Developing a new PI-RADS v2-based nomogram for forecasting high-grade prostate cancer. *Clin Radiol* 2017, 72:458-464.

22. Shi Y, Au JS-K, Thongprasert S, et al. 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* 2014, 9:154-162.
23. Sun P-L, Seol H, Lee HJ, et al. High incidence of EGFR mutations in Korean men smokers with no intratumoral heterogeneity of lung adenocarcinomas: correlation with histologic subtypes, EGFR/TTF-1 expressions, and clinical features. *J Thorac Oncol* 2012, 7:323-330.
24. Sekine A, Tamura K, Satoh H, et al. Prevalence of underlying lung disease in smokers with epidermal growth factor receptor-mutant lung cancer. *Oncol Rep* 2013, 29:2005-2010.
25. Song Z, Zhu H, Guo Z, et al. Correlation of EGFR mutation and predominant histologic subtype according to the new lung adenocarcinoma classification in Chinese patients. *Med Oncol* 2013, 30:645.
26. Shigematsu H, Gazdar AF. Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. *Int J Cancer* 2006, 118:257-262.
27. Motoi N, Szoke J, Riely GJ, et al. Lung adenocarcinoma: modification of the 2004 WHO mixed subtype to include the major histologic subtype suggests correlations between papillary and micropapillary adenocarcinoma subtypes, EGFR mutations and gene expression analysis. *Am J Surg Pathol* 2008, 32:810-827.
28. Rosell R, Moran T, Queralt C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. *N Engl J Med* 2009, 361:958-967.
29. Tanaka T, Matsuoka M, Sutani A, et al. Frequency of and variables associated with the EGFR mutation and its subtypes. *Int J Cancer* 2010, 126:651-655.
30. Gorelik E, Landsittel DP, Marrangoni AM, et al. Multiplexed immunobead-based cytokine profiling for early detection of ovarian cancer. *Cancer Epidemiol Biomarkers Prev* 2005, 14:981-987.
31. Girard N, Sima CS, Jackman DM, et al. Nomogram to predict the presence of EGFR activating mutation in lung adenocarcinoma. *Eur Respir J* 2012, 39:366-372.
32. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015, 350:g7594.
33. Localio AR, Goodman S. Beyond the usual prediction accuracy metrics: reporting results for clinical decision making. *Ann Intern Med* 2012, 157:294-295.
34. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. *Med Decis Making* 2015, 35:162-169.
35. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak* 2008, 8:53.
36. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006, 26:565-574.
37. Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: more than meets the eye. *Lancet Oncol* 2015, 16:e173-e180.

Tables

Table 1. Univariate and multivariate logistic regression models in the development cohort.

Characteristic	EGFR mutant (n=159)	EGFR wild-type (n=143)	Univariate analysis		Multivariate analysis	
			OR (95%CI)	*P* value	OR (95%CI)	*P* value
Age (years)						
Mean ± SD	57.55±9.31	58.43±9.78	0.99 (0.97~1.01)	0.421	NA	
Sex						
Male	58 (36.5)	90 (62.9)	Reference	0.001	NA	
Female	101 (63.5)	53 (37.1)	2.96 (1.85~4.72)			
Smoking history*						
Yes	23 (14.5)	58 (40.6)	Reference	0.001	Reference	
No	136 (85.5)	85 (59.4)	0.25 (0.14~0.43)		0.2 (0.1, 0.4)	<0.001
CEA (μg/L)						
Normal	44 (27.7)	48 (33.6)	Reference	0.267	NA	
High	115 (72.3)	95 (66.4)	1.32 (0.81, 2.16)			
Distribution						
Central	38 (23.9)	40 (28.0)	Reference	0.419	NA	
Peripheral	121 (76.1)	103 (72.0)	1.24 (0.74~2.07)			
Lobe location						
Right upper	45 (28.3)	41 (28.7)	Reference	0.217	NA	
Right middle	19 (11.9)	8 (5.6)	2.16 (0.86~ 5.47)			
Right lower	37 (23.3)	30 (21.0)	1.12 (0.59~ 2.13)			

Left upper	34 (21.4)	32 (22.4)	0.97 (0.51~1.84)		
Left lower	24 (15.1)	32 (22.4)	0.68 (0.35~1.35)		
Long-axis diameter	3.91±1.83	5.17±2.55	0.77 (0.69~0.86)	0.001	NA
Short-axis diameter	3.12±3.07	3.62±1.76	0.89 (0.78~1.02)	0.098	NA
Size					
≤ 3 cm	61 (38.4)	32 (22.4)	Reference	0.003	NA
3 cm	98 (61.6)	111 (77.6)	0.46 (0.28~0.77)		
Contour					
Regular	68 (42.8)	38 (26.6)	Reference	0.003	NA
Irregular	91 (57.2)	105 (73.4)	0.48 (0.30~0.79)		
Lobulation					
≤ 3	75 (47.2)	52 (36.4)	Reference	0.058	NA
3	84 (52.8)	91 (63.6)	0.64 (0.40~1.02)		
Spiculation					
Yes	120 (75.5)	80 (55.9)	2.42 (1.49~3.95)	0.001	NA
No	39 (24.5)	63 (44.1)	Reference		
Texture					
Non-solid	36 (22.6)	27 (18.9)	Reference	0.422	NA
Solid	123 (77.4)	116 (81.1)	0.80 (0.45~1.39)		
Air bronchogram					

Yes	104 (65.4)	69 (48.3)	2.03 (1.28~3.22)	0.003	NA	
No	55 (34.6)	74 (51.7)	Reference			
Bubblelike lucency						
Yes	103 (64.8)	57 (39.9)	2.78 (1.74~4.43)	0.001	2.2 (1.3, 3.8)	0.003
No	56 (35.2)	86 (60.1)	Reference		Reference	
Calcification						
Yes	24 (15.1)	36 (25.2)	0.53 (0.30~0.94)	0.029	NA	
No	135 (84.9)	107 (74.8)	Reference			
Vascular convergence						
Yes	139 (87.4)	101 (70.6)	2.89 (1.60~5.22)	0.001	NA	
No	20 (12.6)	42 (29.4)	Reference			
Fissure attachment						
Yes	50 (31.4)	50 (35.0)	0.85 (0.53~1.38)	0.517	NA	
No	109 (68.6)	93 (65.0)	Reference			
Pleural attachment						
Yes	42 (26.4)	71 (49.7)	0.36 (0.22~0.59)	0.001	0.4 (0.2, 0.7)	0.001
No	117 (73.6)	72 (50.3)	Reference		Reference	
TABD						
Yes	101 (63.5)	42 (29.4)	4.19 (2.58~6.79)	0.001	3.1 (1.8, 5.3)	0.001
No	58 (36.5)	101 (70.6)	Reference			
Peripheral emphysema						
Yes	131 (82.4)	123 (86.0)	0.76	0.390	NA	

			(0.41~1.42)		
No	28 (17.6)	20 (14.0)	Reference		
Peripheral fibrosis					
Yes	127 (79.9)	122 (85.3)	0.68 (0.37~1.25)	0.215	NA
No	32 (20.1)	21 (14.7)	Reference		
Inflammation					
Yes	110 (69.2)	101 (70.6)	0.93 (0.57~ 1.53)	0.784	NA
No	49 (30.8)	42 (29.4)	Reference		
Pleural effusion					
Yes	41 (25.8)	53 (37.1)	0.59 (0.36~0.96)	0.035	NA
No	118 (74.2)	90 (62.9)	Reference		

CEA, carcinoembryonic antigen; EGFR, epidermal growth factor receptor; NA, not applicable; SD, standard deviation; TABD, Thickened adjacent bronchovascular bundles.

* Smoking history is defined as follows: Yes, former and current smokers; No, never smoked.

Table 2. The AUCs of the ROC curves for the nomogram and variables from the logistic regression model in the development and validation cohorts.

	Development cohort			Validation cohort		
	AUC (95%CI)	Sensitivity	Specificity	AUC (95%CI)	Sensitivity	Specificity
Nomogram variables	0.784 (0.733~0.835)	0.679	0.762	0.740 (0.643~0.838)	0.588	0.840
Smoking history	0.630 (0.582~0.679)	0.855	0.406	NA	NA	NA
Bubble-like lucency	0.625 (0.570~0.679)	0.648	0.601	NA	NA	NA
Pleural attachment	0.616 (0.563~0.670)	0.736	0.497	NA	NA	NA
TABD	0.671 (0.618~0.724)	0.635	0.706	NA	NA	NA

AUC, area under the curve; CI, confidence interval; ROC, receiver operating characteristic; TABD, Thickened adjacent bronchovascular bundles.

Figures

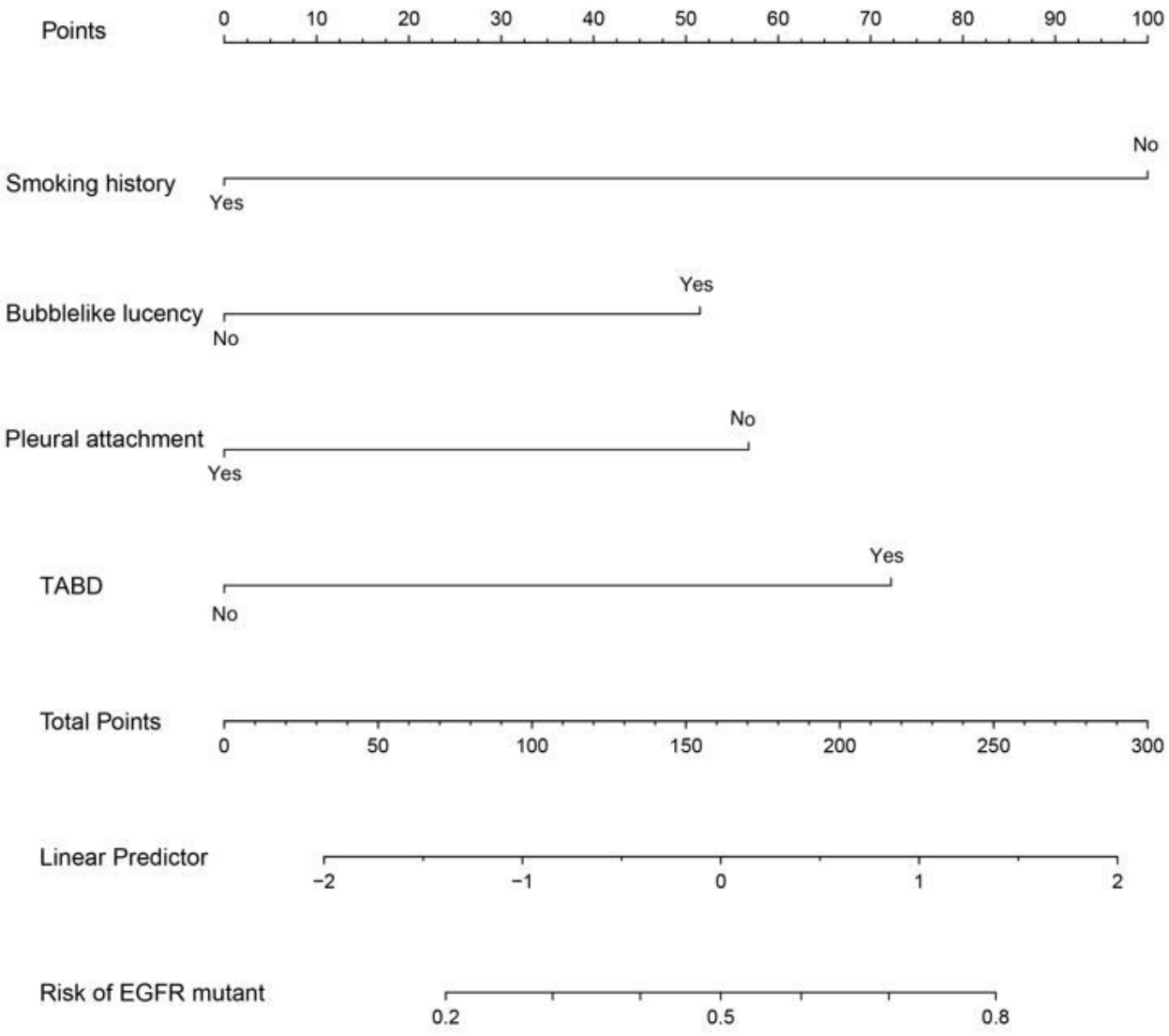

Figure 1

Nomogram for prediction of EGFR mutation status risk and its predictive performance. The nomogram was constructed in the development cohort combining the clinical characteristic, including the smoking history, and CT features, including bubble-like lucency, pleural attachment, and TABD. CT, computed tomography; EGFR, epidermal growth factor receptor; TABD, thickened adjacent bronchovascular bundles.

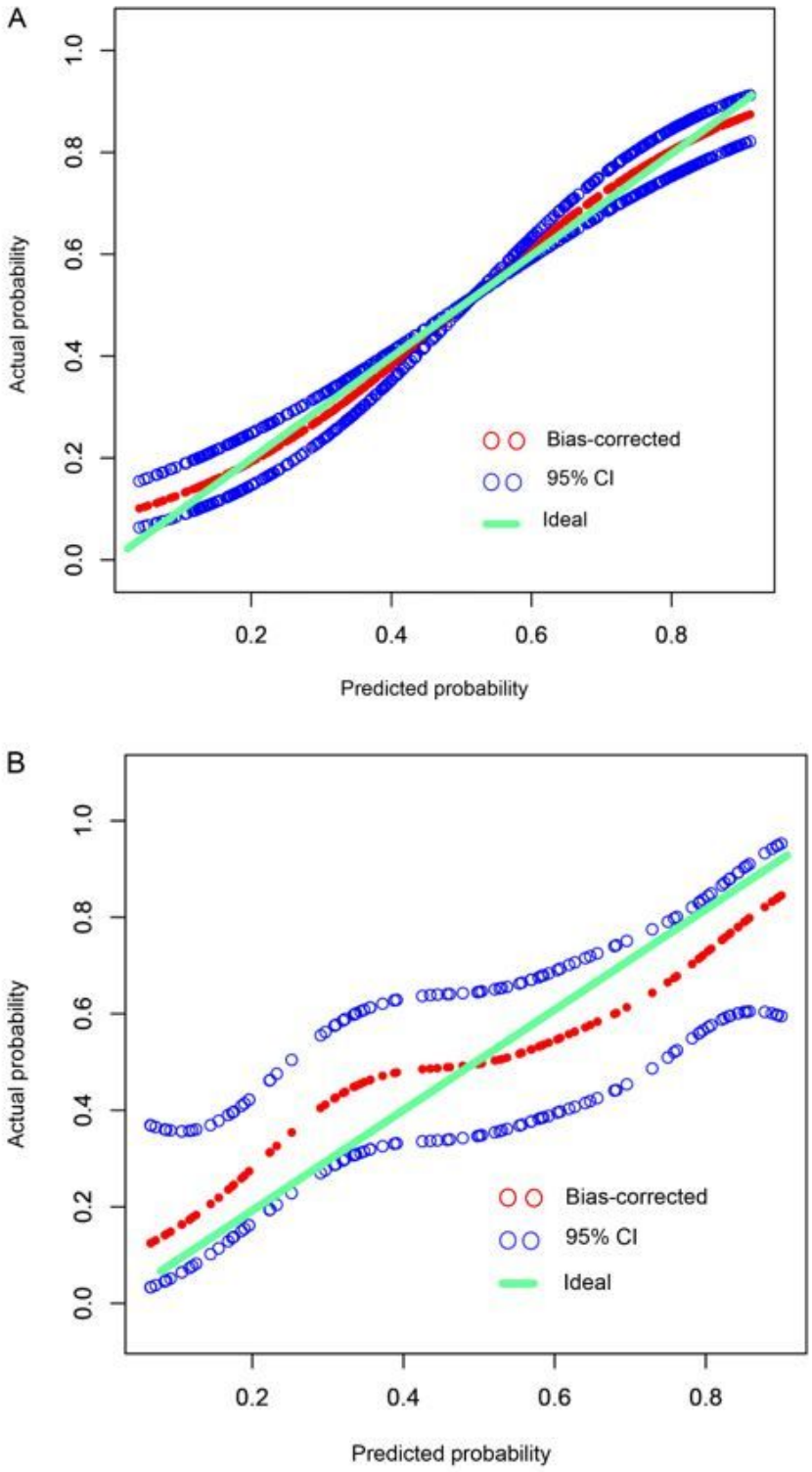

Figure 2

Calibration curve of the nomogram in development (a) and validation (b) cohorts. The x-axis represents the predicted EGFR mutation risk. The y-axis represents the actual EGFR mutation rate. The green line represents a perfect estimated mutation rate by an ideal model. The red circle represents the performance of the nomogram, in which a closer fit to the green line represents a better prediction. The blue circle represents the 95% confidence interval. EGFR, epidermal growth factor receptor.

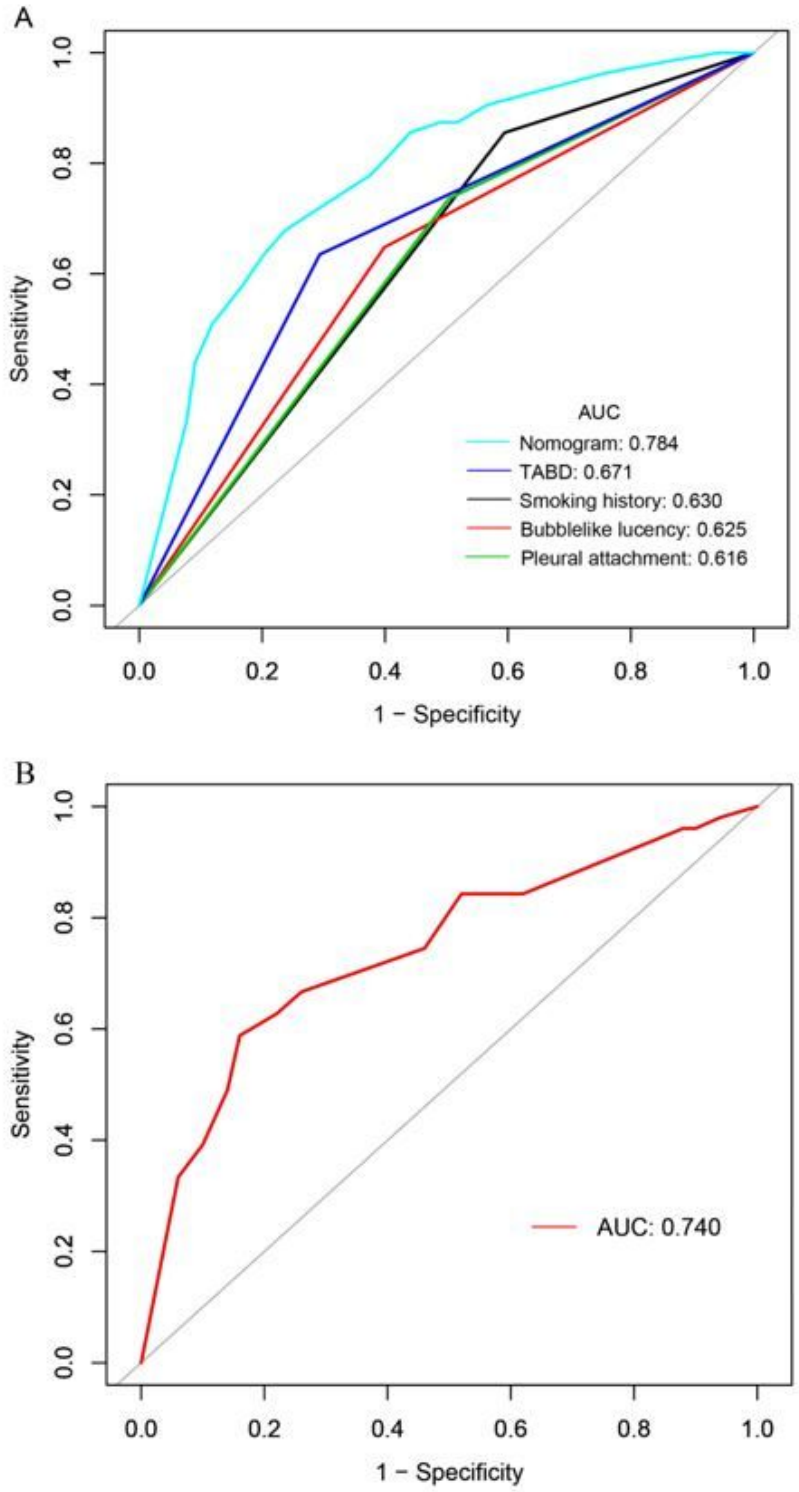

Figure 3

ROC curves of the predictive EGFR mutation status in the development (a) and validation (b) cohorts. The AUC of the hybrid nomogram in the development cohort was 0.786 (Light blue line, Fig. 4a), and the AUC in the validation cohort was 0.740 (Red line, Fig. 4b). AUC, area under the curve; EGFR, epidermal growth factor receptor; ROC, receiver operating characteristic; TABD, thickened adjacent bronchovascular bundles.

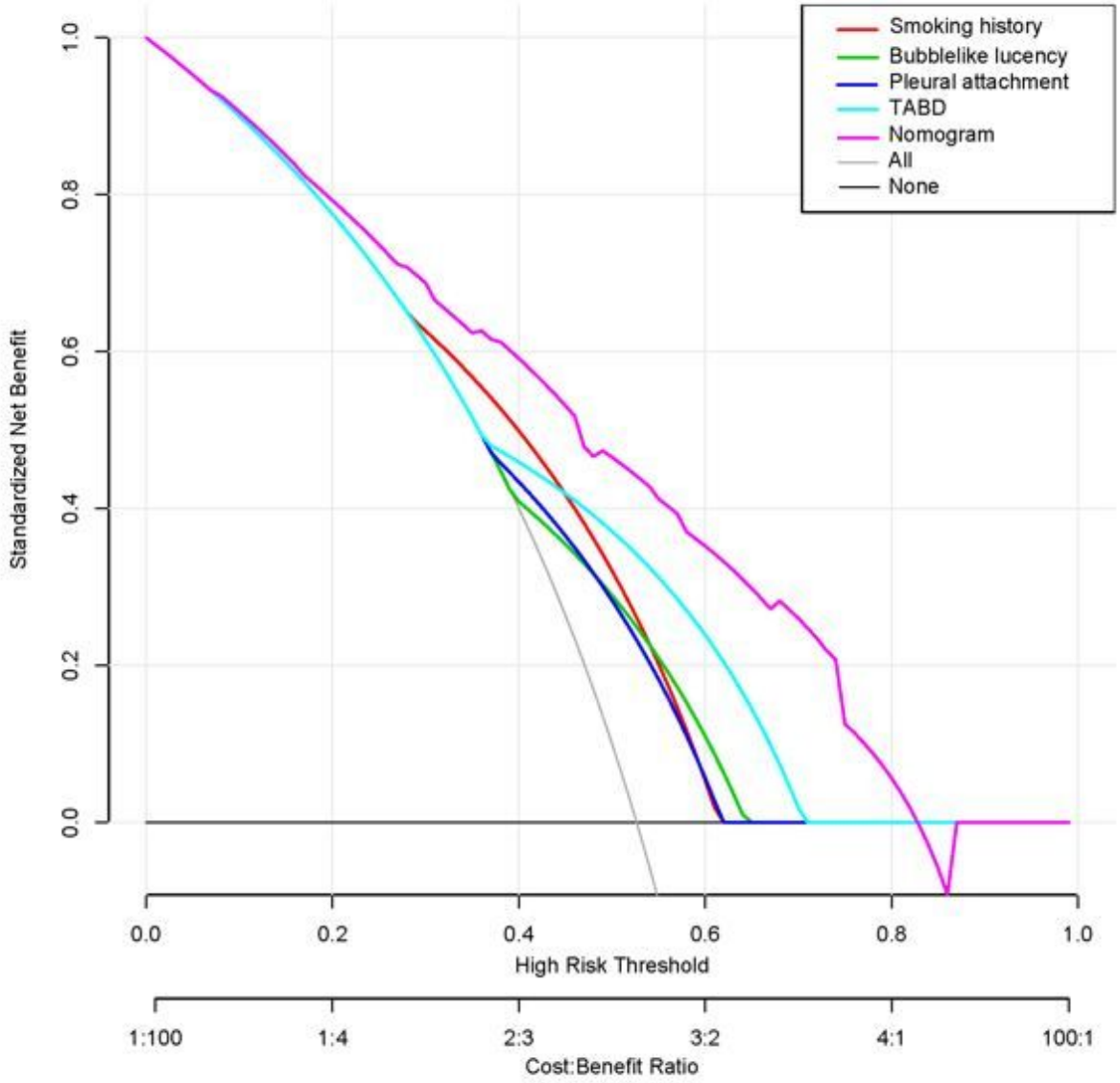

Figure 4

Decision curves of the four risk factors and the hybrid nomogram. The y-axis measures the net benefit. The purple line represents the hybrid nomogram. Red, light blue, dark blue, and green lines represent smoking history, TABD, pleural attachment and bubble-like lucency, respectively. The gray line represents the hypothesis that all patients have EGFR mutations. The black line represents the hypothesis that no patients have EGFR mutations. When the threshold probability is <83%, the hybrid nomogram shows the highest net benefit. EGFR, epidermal growth factor receptor; TABD, Thickened adjacent bronchovascular bundles.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx