Heliyon 10 (2024) e24372

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Dual-energy CT-based radiomics in predicting EGFR mutation status non-invasively in lung adenocarcinoma

Jing-Wen Ma [a,b,1], Xu Jiang [a,1], Yan-Mei Wang [c], Jiu-Ming Jiang [a], Lei Miao [a], Lin-Lin Qi [a], Jia-Xing Zhang [a], Xin Wen [a], Jian-Wei Li [a], Meng Li [a,*,2], Li Zhang [a,**,2]

[a] *Department of Diagnostic Radiology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China*

[b] *Department of Radiology, State Key Laboratory of Cardiovascular Disease, National Clinical Research Center for Cardiovascular Diseases, Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, #167 Bei-Li-Shi Street, Beijing 100037, China*

[c] *GE Healthcare China, Pudong New Town, Shanghai, China*

A R T I C L E I N F O

Keywords:
CT-based radiomics
Dual-energy spectral CT
EGFR mutation
Lung adenocarcinoma
Nomogram

A B S T R A C T

Objectives: Patients with epidermal growth factor receptor (EGFR) mutations in lung adenocarcinoma (LUAD) can benefit from individualized targeted therapy. This study aims to develop, compare, analyse prediction models based on dual-energy spectral computed tomography (DESCT) and CT-based radiomic features to non-invasively predict EGFR mutation status in LUAD.

Materials and methods: Patients with LUAD (n = 175), including 111 patients with and 64 patients without EGFR mutations, were enrolled in the current study. All patients were randomly divided into a training dataset (122 cases) and validation dataset (53 cases) at a ratio of 7:3. After extracting CT-based radiomic, DESCT and clinical features, we built seven prediction models and a nomogram of the best prediction. Receiver operating characteristic (ROC) curves and the mean area under the curve (AUC) values were used for comparisons among the models to obtain the best prediction model for predicting EGFR mutations.

Results: The best distinguishing ability is the combined model incorporating radiomic, DESCT and clinical features for predicting the EGFR mutation status with an AUC of 0.86 (95 % CI: 0.79–0.92) in the training group and an AUC value of 0.83 (95 % CI: 0.73, 0.96) in the validation group.

Conclusions: Our study provides a predictive nomogram non-invasively with a combination of CT-based radiomic, DESCT and clinical features, which can provide image-based biological information for targeted therapy of LUAD with EGFR mutations.

1. Introduction

Lung cancer remains a major threat to human health worldwide [1,2]. Lung adenocarcinoma (LUAD), accounting for more than 50

* Corresponding author.
** Corresponding author.
E-mail addresses: lmcams@163.com (M. Li), zhangli_cicams@163.com (L. Zhang).

[1] Jing-Wen Ma and Xu Jiang contributed equally to this work as co-first authors.

[2] Meng Li and Li Zhang have contributed equally to this work and share corresponding.

https://doi.org/10.1016/j.heliyon.2024.e24372
Received 26 June 2023; Received in revised form 15 December 2023; Accepted 8 January 2024
Available online 12 January 2024

% of lung cancer, is the most common pathological subtype of lung cancer, and is closely related to oncogene mutations [2–4]. Epidermal growth factor receptor (EGFR) is the most common mutation type in LUAD, accounting for 40%–55 % of Asian and 5%–15 % of white patients and is commonly found in female non-smokers [5–9]. LUAD patients with EGFR mutations respond well to tyrosine kinase inhibitors (TKIs), such as gefitinib with an effective rate of over 80 % [10–12]. Therefore, it is crucial to identify EGFR mutations before targeted therapy for LUAD patients.

Although pathological diagnosis is still considered the "gold standard" for detecting gene mutations in cancer lesions, not all lesions are suitable for gene testing through biopsy or surgery in clinical practice. Moreover, due to the heterogeneity of tumous (including intratumoral and intertumoral heterogeneity), tissue specimens obtained by puncture or surgery may not reflect all information on the phenotypes and variations of mutation genes in lesions [13–15]. Liquid biopsy can obtain mutant genes in plasma, but false-negative results may occur if there is not enough DNA shed into the circulation [16]. Imaging techniques have many advantages, such as non-invasiveness, repeatability, convenience, rapidity, and cost-effectiveness, and allow observing the overall lesion. Conventional images can reflect the molecular and microscopic genetic changes of tumors to some extent [6], but conventional imaging features exist subjective and semi-quantitative nature. Dual-energy spectral computed tomography (DESCT) can not only perform routine analysis on lesions but also extract multiple quantitative parameters from images, providing more information for predicting gene mutations in lung cancer [17]. However, there are few studies on DESCT for LUAD EGFR mutations at present. Radiomics is an imaging research method that has received considerable attention in recent years. It uses multiple quantitative features extracted from conventional images for high-throughput analysis and screening to objectively reflect the microscopic biological information of tumors [18]. There are more studies on predicting LUAD EGFR mutations using radiomics; however, most of them lack uniformity for the scanning parameters, which reduces credibility and increases the instability factors of radiomic features, and only focus on radiomic information [19].

Therefore, our study, which is a retrospective study based on a prospective collection, used DESCT images with consistent and standardized scanning parameters to extract DESCT, radiomics features and combined them to construct a prediction model and the best model nomogram to predict EGFR mutation status to assist clinical practice in LUAD.

2. Material and methods

2.1. Clinical data

A retrospective analysis was conducted on the clinical and imaging data of 501 patients with pathologically confirmed lung adenocarcinoma who underwent DESCT examination at the Cancer Hospital of the Chinese Academy of Medical Sciences from May 2013 to December 2015. The inclusion criteria were as follows: (1) patients with LUAD pathology; (2) complete thin-layer (1.25 mm) DESCT images and clinical data stored in the Picture Archiving and Communication Systems (PACS); (3) no history of other malignant tumors; and (4) no previous lung cancer-related treatment, such as chemotherapy, radiotherapy, or immunotherapy. The exclusion criteria were as follows: (1) no EGFR mutation gene testing; (2) non-LUAD histological subtypes; (3) poor image quality that did not meet diagnostic criteria; or (4) tumor boundaries that were difficult to distinguish due to a large amount of pleural effusion or inflammation. This retrospective study was approved by the ethics committee of the Cancer Hospital of the Chinese Academy of Medical Sciences.

2.2. DESCT examination

Chest scans were performed using a GE Healthcare Discovery CT750 HDCT scanner, covering the range from the lung apex to the level of the lower edge of the adrenal gland. A Gemstone spectral imaging (GSI) mode protocol was used. The scan parameters were as follows: tube current of 550 mA, tube rotation time of 0.6 s, and collimator width of 40 mm. Before scanning, contrast agent (Ultravist 300; Bayer AG, Germany) was injected intravenously at a rate of 2.5 ml/s with a volume of 85–100 ml. The enhanced scan was performed 35 s after contrast agent injection.

2.3. DESCT image segmentation and feature extraction

Image segmentation and analysis were performed using the GSI Volume Viewer on the post-processing workstation (Advantage Workstation 4.6, GE Healthcare, Milwaukee, WI, USA) [20,21]. One radiologist with 8 years of experience in image diagnosis selected the axial images, manually located the lesion centre, and outlined the region of interest (ROI), with an ROI area no less than two-thirds of the lesion and avoiding cavities, air pockets, calcification, vessels, and lung collapse. DESCT features were automatically extracted by the software.

2.4. Radiomics image segmentation and feature extraction

Preoperative thin-layer CT images of patients were imported into ITK-SNAP 3.8 software (http://www.itksnap.org) in Digital Imaging and Communications in Medicine (DICOM) format. The radiologist with 8 years of experience in image diagnosis outlined the ROI on the axial images and confirmed and repaired it on the coronal and sagittal images to generate the volume of interest (VOI). Another radiologist with 15 years of experience in image diagnosis reviewed the VOI of each LUAD. Both radiologists were blinded to all clinical and pathological data of patients. Radiomic features were extracted using the Python Package Py-radiomics (http://

pyradiomics.readthedocs.io/en/Latest/).

2.5. Clinical, DESCT, and radiomics feature selection

Three clinical features were analyzed in the study including smoking history, sex, and age. Among which, the features with significant statistical difference were ultimately included in the model by using univariate analysis.

Z scores were used to standardize the radiomics and DESCT features. The Spearman correlation coefficient was used to eliminate redundant radiomics features, with a truncation coefficient of 0.90. The LASSO algorithm was applied to select the final radiomic features. For the DESCT model, max-relevance and min-redundancy (MRMR) were used to select the most five representative features.

2.6. Prediction model construction and validation

The logistic regression (LR), support vector machine (SVM), random forest (RF), Bayes and adaptive boosting (AdaBoost) classification methods were used to build models based on extracted clinical, radiomic and DESCT features. LR was used to establish the clinical feature model (M_C). SVM had better classification performance than other classifiers for radiomics and DESCT features (All results of all the classifiers in the different prediction models can be found in Supplementary Materials Table 1). Then, based on the Rad-scores calculated for M_C, M_D, and M_R, different combination models were established using multivariate logistic regression, including the clinical and radiomics feature combined model ($M_{C\text{-}R}$), clinical and DESCT feature combined model ($M_{C\text{-}D}$), radiomics and DESCT feature combined model ($M_{R\text{-}D}$), and clinical, DESCT, and radiomics feature three combined model ($M_{C\text{-}R\text{-}D}$). These seven models were validated in the validation group, and a nomogram was constructed based on the best prediction model. The predictive efficacy of each model for the LUAD EGFR mutation status was evaluated by the receiver operating characteristic (ROC) curve and the area under the curve (AUC). The decision curve was used to evaluate the clinical utility of the models.

2.7. Statistical analysis and model evaluation

Python 3.8.1 were used for statistical analysis. The chi-square test was used for comparison of categorical variables, while the Wilcoxon signed-rank test or *t*-test was used for comparison of continuous variables. Multivariate logistic regression and LASSO analysis were performed using the scikit-learn toolbox. The ROC curve was generated using the Matplotlib toolbox, and AUC, sensitivity, specificity, and accuracy were calculated to evaluate the predictive performance of different models. Finally, the decision curve and the best model nomogram were plotted. The DeLong test was used to compare the AUC values of different models, with $P < 0.05$ indicating statistical significance.

3. Results

3.1. Selection of clinical features

A total of 175 patients who met the inclusion criteria, including 81 males and 94 females, with an age range of 30–76 (mean age ± SD: 57.68 ± 9.72) years, were enrolled. 111 patients were EGFR mutation-positive, and 64 were the wild-type. No significant differences were observed in sex, smoking history and age ($P = 0.86$, 0.12 and 0.28, respectively) between the training group and the validation group (Table 1). Significant differences in sex and smoking history were observed between patients with wild-type EGFR and patients with EGFR mutation in the training group (both $P = 0.003$, Table 2). In the validation group, only the difference in smoking history between patients with wild-type EGFR and patients with EGFR mutation was statistically significant ($P = 0.008$) while sex showed a marginally significant difference between the two groups ($P = 0.05$) (Table 3).

3.2. Selection of DESCT and radiomics features

A total of 24 DESCT features, including CT values (HU) at each energy level of the lesion ROI in the range of 40–140 keV (with a 10 keV interval), substance concentrations such as water concentration and iodine concentration, and effective atomic number, were extracted. The final DESCT model included five DESCT features, namely, water HAP, calcium water, water calcium, HAP iodine, and

Table 1
Comparison and analysis of the clinical data of patients.

Clinical characteristic	Number	Training Group (n = 122)	Validation Group (n = 53)	P value
Sex [n (%)]				
Male	81	57 (46.7)	24 (45.3)	0.86
Female	94	65 (53.3)	29 (54.7)	
Smoking history [n (%)]				
No	107	70 (57.4)	37 (69.8)	0.12
Yes	68	52 (42.6)	16 (30.2)	
Age	175	58.50 (52.00, 65.00)*	57.00 (49.70, 65.30)*	0.28

*Median (Interquartile range). P < 0.05 has significance differences.

Table 2
Comparative analysis of clinical features of patients with EGFR gene mutation or wild-type EGFR in the training group.

Clinical characteristic	Number	Wild-type EGFR (n = 45)	EGFR mutation (n = 77)	P value
Sex [n (%)]				
Male	57	29 (64.4)	28 (36.4)	0.003
Female	65	16 (35.6)	49 (63.6)	
Smoking history [n (%)]				
No	70	18 (40.0)	52 (67.5)	0.003
Yes	52	27 (60.0)	25 (32.5)	
Age	122	58.40 ± 9.98*	58.19 ± 9.25*	0.91

*Mean age ± standard deviation. P < 0.05 indicates significant differences.

Table 3
Comparative analysis of clinical features of EGFR gene mutant patients and wild-type patients in the validation group.

Clinical characteristic	Number	Wild-type EGFR (n = 19)	EGFR mutation (n = 34)	P value
Sex [n (%)]				
Male	24	12 (63.2)	12 (35.3)	0.05
Female	29	7 (36.8)	22 (64.7)	
Smoking history [n (%)]				
No	37	9 (47.3)	28 (82.4)	0.008
Yes	16	10 (52.6)	6 (17.7)	
Age	53	55.68 ± 10.22*	56.68 ± 10.47*	0.74

*Mean age ± standard deviation. P < 0.05 indicates significant differences.

water iodine. The extracted 1316 radiomic features were categorized into five groups: (1) First-order features including 18 intensity statistics and 14 shape features; (2) 75 multi-dimensional texture features including 24 Gy Level Co-occurrence Matrix (GLCM), 16 Gy Level Size Zone Matrix (GLSZM), 16 Gy Level Run Length Matrix (GLRLM), 14 Gy Level Dependence Matrix (GLDM) and 5 Neighboring Grey Tone Difference Matrix (NGTDM) Features; (3)1209 Transformed first-order and textural features including: 744 wavelet-decomposed features in frequency channels LHL, LLH, HHH, HLH, HLL,HHL, LHH and LLL; 186 Laplacian of Gaussian (LoG) filtered features with sigma of 2.0 mm and 3.0 mm; 279 3D local binary pattern (LBP) filtered texture features with 2-level spherical harmonics and icosphere with radius 1 and 1 subdivision. After removal of redundancies, the final model included five radiomics features, including one first-order feature, two GLCM features, one GLRLM feature, and one GLSZM feature.

3.3. Establishment and performance of the prediction model

We constructed a total of seven predictive models in the training group, namely, M_C, M_D, M_R, $M_{C\text{-}R}$, $M_{C\text{-}D}$, $M_{R\text{-}D}$, and $M_{C\text{-}R\text{-}D}$, and plotted the ROC curves for all seven models (Fig. 1). The performance of each model was evaluated based on the ROC curve, and the results showed that $M_{C\text{-}R\text{-}D}$ had the highest AUC value in both the training and validation groups, with an AUC value of 0.86, an accuracy of 0.81, and a specificity of 0.89, and a sensitivity of 0.77 in the training group. Similarly, in the validation group, $M_{C\text{-}R\text{-}D}$ had an AUC value of 0.83, with an accuracy, sensitivity, and specificity of 0.79, 0.82, and 0.74, respectively. A nomogram of $M_{C\text{-}R\text{-}D}$ was generated in the training group (Fig. 2), which can be used for personalized prediction of LUAD EGFR mutations. The higher the calculated score was, the higher the probability of EGFR mutations in patients would be. The ROC curves, AUC values, accuracy, sensitivity, and specificity for the seven models in the training and validation groups are shown in Tables 4 and 5. The AUC differences

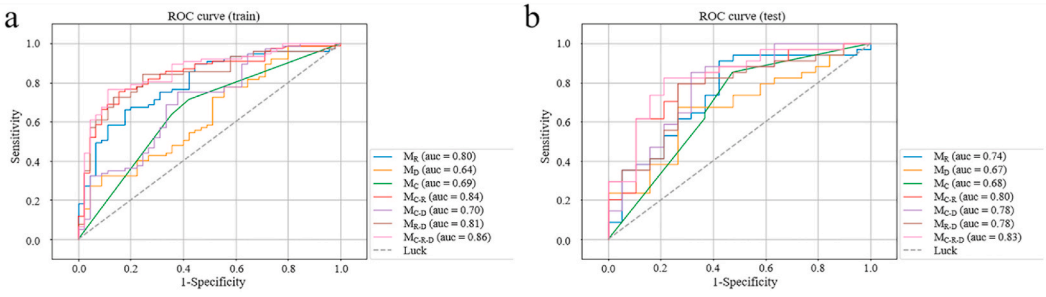

Fig. 1. Two receiver operating characteristic (ROC) curves was plotted to evaluate the performance of radiomics, DESCT, and four combined models for predicting EGFR mutation status in LUAD in training group and validation group, respectively.
a. ROC curve for seven models in the training group; b. ROC curve for seven models in the validation group. M_C, clinical model; M_R, radiomics model; M_D, DESCT model; $M_{C\text{-}R}$, clinical-radiomics feature combined model; $M_{R\text{-}D}$, radiomics and DESCT feature combined model; $M_{C\text{-}D}$, clinical and DESCT feature combined model; $M_{C\text{-}R\text{-}D}$, clinical, DESCT, and radiomics feature combined model.

between the seven models in the validation group are shown in Supplementary Materials Table 2. Decision curve analysis showed that the predictive performance of $M_{C\text{-}R\text{-}D}$ in the training and validation groups was superior to that of the other models (Fig. 3).

4. Discussion

This study evaluated the feasibility of prediction models based on clinical, DESCT, and radiomic features to predict the LUAD EGFR mutation status. All data in this study were obtained from the consistent CT scanner with identical scanning parameters. The results showed that the combined model based on these three features was the optimal model for predicting LUAD EGFR mutations, and the model's nomogram visualized the prediction results and could serve as a clinical tool for predicting LUAD EGFR mutation.

Clinical model (M_C) was developed consisting of smoking history and sex. In the training and validation groups, the AUCs of this model were 0.69 (95 % CI: 0.60–0.77) and 0.68 (95 % CI: 0.54–0.83), respectively. Numerous epidemiological studies have shown that the incidence of EGFR mutations is higher in Asian populations, women, and non-smoking lung adenocarcinoma patients [5,7–9]. The result of this study is consistent with previous reports. However, the prediction model based on smoking history and sex had low diagnostic efficiency and could not meet clinical diagnostic requirement.

In recent years, DESCT has become a hot topic in quantitative imaging research and its application in tumor assessment [22]. The energy spectrum scanner used in this study employs a single-source 80 keV and 140 keV tube voltage fast switching for scanning, which simultaneously obtains single-energy level images, corresponding energy spectrum curves, different material concentration parameters, effective atomic numbers, and other objective and quantitative parameters [17,22]. This is conducive to understanding the histopathological information of the lesions [23]. In this study, the predictive performance of M_D based on five DESCT features for LUAD EGFR mutations was not satisfactory, with an AUC of 0.64 (95 % CI: 0.54–0.72) in the training group and 0.67 (95 % CI: 0.53–0.83) in the validation group, which was lower than that of M_C. However, the predictive performance of $M_{C\text{-}D}$, which combined clinical features with M_D, was improved, with an AUC of 0.78 in the validation group, consistent with the findings of Li et al. [20]. Li et al. [20] used standardized iodine concentration values and smoking history to predict EGFR mutations in lung adenocarcinoma, with an AUC value of 0.70, which has practical value. Currently, there is limited research on using DESCT features to predict LUAD EGFR mutation status, and more exploration is needed.

In this study, five radiomics features were included, among which the first-order statistical features primarily describe the distribution of pixel or voxel intensities within the tumor region. The grey-level cooccurrence matrix (GLCM) reflects the spatial relationship of grey levels and thereby describes the texture distribution and characteristics inside the tumor. GLSZM (Grey-Level Size Zone) and GLRLM (Grey-Level Run Length Matrix) are closely related to texture heterogeneity [24] and thus were hypothesized to be used for predicting LUAD EGFR mutations and establishing prediction models. The results showed that the predictive performance of the four models involving radiomics features, namely, M_R, $M_{C\text{-}R}$, $M_{R\text{-}D}$, and $M_{C\text{-}R\text{-}D}$, was good, with AUC values more than or equal to 0.8 in the training group. Moreover, in agreement with previous studies [25–28], combining radiomic features with clinical features improved the predictive performance. Radiomics can extract data information from conventional CT and MR images that cannot be captured by the naked eye, and some scholars believe that radiomics can non-invasively reflect tumor heterogeneity [18]. Tumor heterogeneity has been recognized as closely associated with tissue hypoxia and necrosis, cell proliferation, and angiogenesis, all of which are positively correlated with the invasiveness and poor prognosis of malignant tumors. Its expression is of great importance for studying clinical treatment and tumor drug resistance [29,30]. Although the predictive value of CT radiomics features for predicting

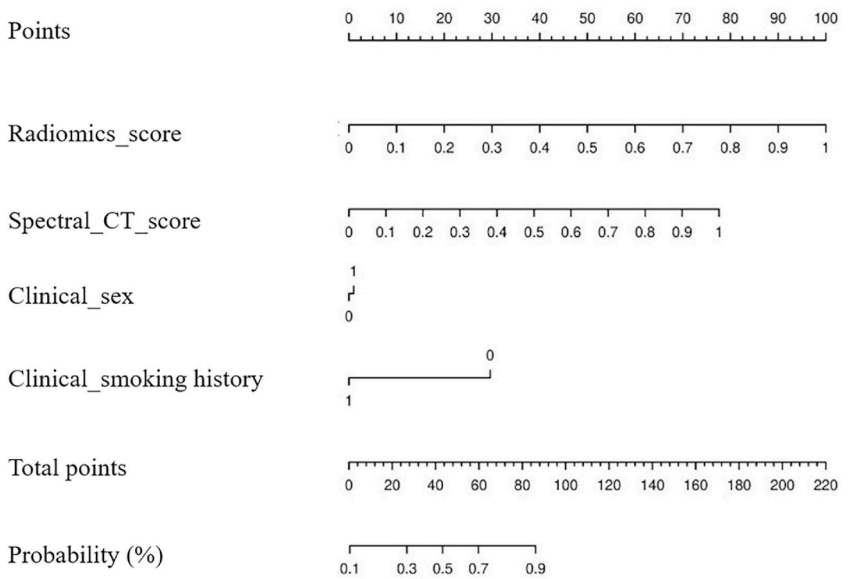

Fig. 2. Clinical-DESCT-radiomics model nomogram in the training cohort.

Table 4
The predictive power of seven prediction models in the training group on the mutation state of EGFR.

Predictive Model	AUC (95 % CI)	Accuracy	Sensitivity	Specificity
M_C	0.69 (0.60, 0.77)	66.40 %	71.40 %	57.80 %
M_R	0.80 (0.72, 0.87)	72.10 %	66.20 %	82.20 %
M_D	0.64 (0.54, 0.72)	54.10 %	32.50 %	91.10 %
$M_{C\text{-}R}$	0.84 (0.77, 0.90)	77.90 %	72.70 %	86.70 %
$M_{R\text{-}D}$	0.81 (0.73, 0.88)	79.50 %	81.80 %	75.60 %
$M_{C\text{-}D}$	0.70 (0.62, 0.78)	70.50 %	75.30 %	62.20 %
$M_{C\text{-}R\text{-}D}$	0.86 (0.79, 0.92)	81.10 %	76.60 %	88.90 %

AUC, area under the curve; CI, confidence interval; M_C, clinical model; M_R, radiomics model; M_D, DESCT model; $M_{C\text{-}R}$, clinical-radiomics feature combined model; $M_{R\text{-}D}$, radiomics and DESCT feature combined model; $M_{C\text{-}D}$, clinical and DESCT feature combined model; $M_{C\text{-}R\text{-}D}$, clinical, DESCT, and radiomics feature combined model.

Table 5
The predictive efficacy of seven predictive models in the validation group on EGFR mutation status.

Predictive Model	AUC (95 % CI)	Accuracy	Sensitivity	Specificity
M_C	0.68 (0.54, 0.83)	64.10 %	64.70 %	63.20 %
M_R	0.74 (0.61, 0.89)	67.90 %	73.50 %	57.90 %
M_D	0.67 (0.53, 0.83)	62.30 %	67.60 %	52.60 %
$M_{C\text{-}R}$	0.80 (0.68, 0.94)	75.50 %	79.40 %	68.40 %
$M_{R\text{-}D}$	0.78 (0.65, 0.91)	77.40 %	79.40 %	73.70 %
$M_{C\text{-}D}$	0.78 (0.66, 0.93)	77.40 %	82.40 %	68.40 %
$M_{C\text{-}R\text{-}D}$	0.83 (0.73, 0.96)	79.20 %	82.40 %	73.70 %

AUC, area under the curve; CI, confidence interval; M_C, clinical model; M_R, radiomics model; M_D, DESCT model; $M_{C\text{-}R}$, clinical-radiomics feature combined model; $M_{R\text{-}D}$, radiomics and DESCT feature combined model; $M_{C\text{-}D}$, clinical and DESCT feature combined model; $M_{C\text{-}R\text{-}D}$, clinical, DESCT, and radiomics feature combined model.

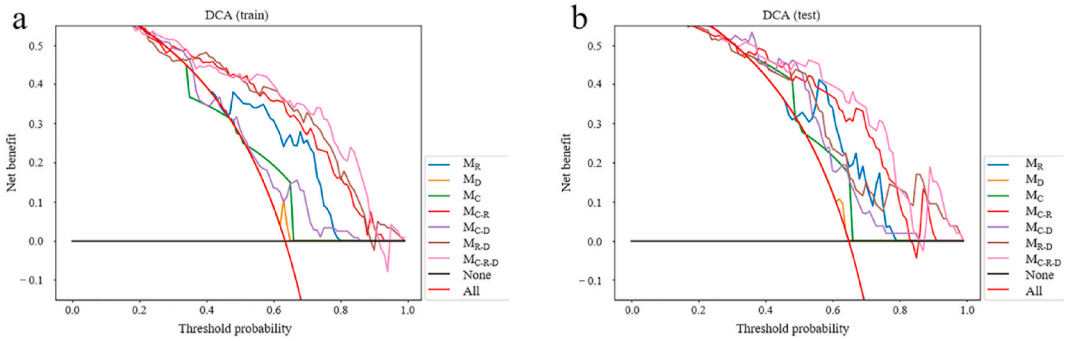

Fig. 3. The decision curves of the seven models: (a) training cohort and (b) validation cohort, respectively. M_C, clinical model; M_R, radiomics model; M_D, DESCT model; $M_{C\text{-}R}$, clinical-radiomics feature combined model; $M_{R\text{-}D}$, radiomics and DESCT feature combined model; $M_{C\text{-}D}$, clinical and DESCT feature combined model; $M_{C\text{-}R\text{-}D}$, clinical, DESCT, and radiomics feature combined model.

LUAD EGFR mutations is lower than that of standard pathological and molecular biological detection, many studies have shown good predictive performance, with most studies having AUCs greater than 0.8 [25,31–34]. However, some studies have shown unsatisfactory prediction results [26,27], which we believe may be related to the inconsistent data parameters used.

In this study, we established four combined models based on clinical, DESCT, and radiomic features. Among the four combined models, $M_{C\text{-}R\text{-}D}$ had the highest AUC value for predicting LUAD EGFR mutations in both the training and validation groups, as well as ideal sensitivity, specificity, and accuracy, indicating that $M_{C\text{-}R\text{-}D}$ based on radiomics, DESCT, and clinical features has the potential to predict whether LUAD has EGFR mutations. Moreover, visualization of the model was achieved by nomogram, and decision curve analysis demonstrated its good clinical application value as a non-invasive, quantitative, convenient, and fast diagnostic method.

Limitations of this study include the following: (1) our data were obtained from a single-centre medical institution, and further multicentre studies are needed to validate the applicability of the results; (2) this was a retrospective study, and all patients were from Asia, so the potential for selection bias could not be avoided; and (3) the sample size of this study was relatively small; thus, larger sample-size studies are needed to explore the generalizability of the results.

5. Conclusions

In summary, this study found that the $M_{C\text{-}R\text{-}D}$ combining clinical, DESCT and radiomic features, can serve as a clinical tool for predicting LUAD EGFR mutations and provide a basis for assisting individualized clinical treatment decisions.

Ethics statement

The study was approved by the Cancer Hospital, Chinese Academy of Medical Sciences ethics commission (NCC2016G-029). Data were anonymized for retrospective analysis and thus ethics commission waived the requirement for informed consent.

Funding

This research was supported by the National Natural Science Foundation of China (81601494) and the Chinese Academy of Medical Science Innovation Fund for Medical Sciences (2021-12M-C&T-B-061).

Data availability statement

Data will be made available on request.

CRediT authorship contribution statement

Jing-Wen Ma: Writing – original draft. **Xu Jiang:** Writing – review & editing. **Yan-Mei Wang:** Methodology, Software. **Jiu-Ming Jiang:** Investigation, Methodology. **Lei Miao:** Data curation. **Lin-Lin Qi:** Data curation. **Jia-Xing Zhang:** Data curation. **Xin Wen:** Data curation. **Jian-Wei Li:** Data curation. **Meng Li:** Conceptualization, Writing – review & editing, Funding acquisition, Supervision. **Li Zhang:** Conceptualization, Writing – review & editing, Funding acquisition, Supervision.

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.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e24372.

Abbreviations

LUAD	lung adenocarcinoma
EGFR	epidermal growth factor receptor
TKIs	tyrosine kinase inhibitors
DESCT	dual-energy spectral computed tomography
ROI	region of interest
VOI	volume of interest
LASSO	least absolute shrinkage and selection operator regression
MRMR	max-relevance and min-redundancy
SVM	support vector machine
RF	random forest
ROC	receiver operating characteristic
AUC	area under the curve
WC	water concentration
IC	iodine concentration
GLCM	grey-level cooccurrence matrix
GLRLM	grey-level run-length matrix
GLSZM	grey-level size-zone matrix
GLDM	grey-level dependence matrix
NGTDM	neighboring grey-tone difference matrix

References

[1] A.A. Thai, et al., Lung cancer, Lancet 398 (10299) (2021) 535–554.
[2] R.L. Siegel, et al., Cancer statistics, 2023, CA Cancer J. Clin. 73 (1) (2023) 17–48.
[3] H. Sung, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (3) (2021) 209–249.
[4] N. Duma, R. Santana-Davila, J.R. Molina, Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment, Mayo Clin. Proc. 94 (8) (2019) 1623–1640.
[5] G.D.C. Santos, F.A. Shepherd, M.S. Tsao, EGFR mutations and lung cancer, Annu. Rev. Pathol. 6 (2011) 49–69.
[6] T. Kohno, et al., Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer, Transl. Lung Cancer Res. 4 (2) (2015) 156–164.
[7] Y. Liu, et al., CT features associated with epidermal growth factor receptor mutation status in patients with lung adenocarcinoma, Radiology 280 (1) (2016) 271–280.
[8] H. Shigematsu, et al., Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers, J. Natl. Cancer Inst. 97 (5) (2005) 339–346.
[9] Y. Yatabe, et al., EGFR mutation testing practices within the Asia Pacific region: results of a multicenter diagnostic survey, J. Thorac. Oncol. 10 (3) (2015) 438–445.
[10] T.J. Lynch, et al., Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib, N. Engl. J. Med. 350 (21) (2004) 2129–2139.
[11] J.G. Paez, et al., EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy, Science 304 (5676) (2004) 1497–1500.
[12] W. Pao, et al., EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib, Proc. Natl. Acad. Sci. U.S.A. 101 (36) (2004) 13306–13311.
[13] N. McGranahan, C. Swanton, Clonal heterogeneity and tumor evolution: past, present, and the future, Cell 168 (4) (2017) 613–628.
[14] J.H. Rasmussen, et al., Intratumor heterogeneity of PD-L1 expression in head and neck squamous cell carcinoma, Br. J. Cancer 120 (10) (2019) 1003–1006.
[15] J.H. Rasmussen, et al., Intratumor heterogeneity is biomarker specific and challenges the association with heterogeneity in multimodal functional imaging in head and neck squamous cell carcinoma, Eur. J. Radiol. 139 (2021) 109668.
[16] J.K.R. Nair, et al., Radiogenomic models using machine learning techniques to predict EGFR mutations in non-small cell lung cancer, Can. Assoc. Radiol. J. 72 (1) (2021) 109–119.
[17] C. Kim, et al., Application of dual-energy spectral computed tomography to thoracic oncology imaging, Korean J. Radiol. 21 (7) (2020) 838–850.
[18] P. Lambin, et al., Radiomics: extracting more information from medical images using advanced feature analysis, Eur. J. Cancer 48 (4) (2012) 441–446.
[19] L. Joo, et al., Stability of MRI radiomic features according to various imaging parameters in fast scanned T2-FLAIR for acute ischemic stroke patients, Sci. Rep. 11 (1) (2021) 17143.
[20] M. Li, et al., Identification of epidermal growth factor receptor mutations in pulmonary adenocarcinoma using dual-energy spectral computed tomography, Eur. Radiol. 29 (6) (2019) 2989–2997.
[21] M. Li, et al., Quantitative features of dual-energy spectral computed tomography for solid lung adenocarcinoma with EGFR and KRAS mutations, and ALK rearrangement: a preliminary study, Transl. Lung Cancer Res. 8 (4) (2019) 401–412.
[22] D. Simons, M. Kachelriess, H.P. Schlemmer, Recent developments of dual-energy CT in oncology, Eur. Radiol. 24 (4) (2014) 930–939.
[23] M. Kaup, et al., Dual-energy computed tomography virtual monoenergetic imaging of lung cancer: assessment of optimal energy levels, J. Comput. Assist. Tomogr. 40 (1) (2016) 80–85.
[24] J.R. Zamacona, et al., Assessing diagnostic complexity: an image feature-based strategy to reduce annotation costs, Comput. Biol. Med. 62 (2015) 294–305.
[25] T.Y. Jia, et al., Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling, Eur. Radiol. 29 (9) (2019) 4742–4750.
[26] Y. Liu, et al., Radiomic features are associated with EGFR mutation status in lung adenocarcinomas, Clin. Lung Cancer 17 (5) (2016) 441–448.e6.
[27] E.R. Velazquez, et al., Somatic mutations drive distinct imaging phenotypes in lung cancer, Cancer Res. 77 (14) (2017) 3922–3930.
[28] J. Choe, et al., CT radiomics-based prediction of anaplastic lymphoma kinase and epidermal growth factor receptor mutations in lung adenocarcinoma, Eur. J. Radiol. 139 (2021) 109710.
[29] B. Ganeshan, et al., Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival, Eur. Radiol. 22 (4) (2012) 796–802.
[30] M. Gerlinger, et al., Intratumor heterogeneity and branched evolution revealed by multiregion sequencing, N. Engl. J. Med. 366 (10) (2012) 883–892.
[31] D. Hong, et al., Radiomics signature as a predictive factor for EGFR mutations in advanced lung adenocarcinoma, Front. Oncol. 10 (2020) 28.
[32] X. Lu, et al., A novel radiomic nomogram for predicting epidermal growth factor receptor mutation in peripheral lung adenocarcinoma, Phys. Med. Biol. 65 (5) (2020) 055012.
[33] B. Zhang, et al., Deep CNN model using CT radiomics feature mapping recognizes EGFR gene mutation status of lung adenocarcinoma, Front. Oncol. 10 (2020) 598721.
[34] G. Zhang, et al., Predicting EGFR mutation status in lung adenocarcinoma: development and validation of a computed tomography-based radiomics signature, Am. J. Cancer Res. 11 (2) (2021) 546–560.