Intratumoral and peritumoral CT-based radiomics strategy reveals distinct subtypes of non-small-cell lung cancer

Xing Tang · Haolin Huang · Peng Du · Lijuan Wang · Hong Yin · Xiaopan Xu

Received: 8 October 2021 / Accepted: 4 April 2022 / Published online: 17 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Purpose To evaluate a new radiomics strategy that incorporates intratumoral and peritumoral features extracted from lung CT images with ensemble learning for pretreatment prediction of lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD).

Methods A total of 105 patients (47 LUSC and 58 LUAD) with pretherapy CT scans were involved in this retrospective study, and were divided into training (n = 73) and testing (n = 32) cohorts. Seven categories of radiomics features involving 3078 metrics in total were extracted from the intra- and peritumoral regions of each patient’s CT data. Student’s t tests in combination with three feature selection methods were adopted for optimal features selection. An ensemble classifier was developed using five common machine learning classifiers with these optimal features. The performance was assessed using both training and testing cohorts, and further compared with that of Visual Geometry Group-16 (VGG-16) deep network for this predictive task.

Results The classification models developed using optimal feature subsets determined from intratumoral region and peritumoral region with the ensemble classifier achieved mean area under the curve (AUC) of 0.87, 0.83 in the training cohort and 0.66, 0.60 in the testing cohort, respectively. The model developed by using the optimal feature subset selected from both intra- and peritumoral regions with the ensemble classifier achieved great performance improvement, with AUC of 0.87 and 0.78 in both cohorts, respectively, which are also superior to that of VGG-16 (AUC of 0.68 in the testing cohort).

Conclusions The proposed new radiomics strategy that extracts image features from the intra- and peritumoral regions with ensemble learning could greatly improve the diagnostic performance for the histological subtype stratification in patients with NSCLC.

Keywords Lung squamous cell carcinoma · Lung adenocarcinoma · CT radiomics · Intra- and peritumoral regions · Ensemble learning

Abbreviations
AUC Area under the curve
CM Co-occurrence matrices
CNN Convolutional neural network
CT Computed tomography
FN False negative
FP False positive
GLCM Gray-level co-occurrence matrix
GLDM Gray-level dependence matrix
GLRLM Gray-level run length matrix
GLSZM Gray-level size zone matrix
LASSO Least absolute shrinkage and selection operator
LBP Local binary pattern
LUAD Lung adenocarcinoma
LUSC Lung squamous cell carcinoma
MID Mutual information difference
mpMRI Multiparametric magnetic resonance imaging
mRMR Minimum redundancy maximum relevance
NGTDM Neighboring gray-tone difference matrix
NSCLC Nonsmall-cell lung cancer

Xing Tang and Haolin Huang have contributed equally to this work, therefore, they are co-first authors.

* Xiaopan Xu
alexander-001@163.com

1 Department of Radiology, Xijing Hospital, Air Force Medical University, No. 169 Changle West Road, Xi’an 710032, Shaanxi, People’s Republic of China

2 School of Biomedical Engineering, Air Force Medical University, Xi’an, Shaanxi, People’s Republic of China
PET-CT  Positron emission tomography computed tomography  
QDA  Quadratic discriminant analysis  
RBF  Radial basis function  
RF  Random forest  
RLM  Run length matrix  
ROC  Receiver-operating characteristic curve  
SVM  Support vector machine  
SVM-RFE  Support vector machine-based recursive feature elimination  
TN  True negative  
TP  True positive  
VGG  Visual geometry group network  
XGBoost  Extreme gradient boosting

Introduction

Lung cancer is the most frequently occurring cancer and the leading cause of cancer-related death in men globally (Sung et al. 2020). In women, lung cancer is the third most commonly diagnosed cancer and the second most leading cause of cancer-related death (Sung et al. 2020). Approximately 85% of primary lung malignancies are non-small-cell lung cancer (NSCLC), and the 5-year survival rate is less than 20% (Bashir et al. 2019; Herbst et al. 2008; Bray, et al. 2018; Su 2019; Ma et al. 2018a).

Lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD) are two major histological subtypes of NSCLC that constitute approximately 35% and 60% of primary NSCLC cases, respectively (Bashir et al. 2019; Herbst et al. 2008; Su 2019; Zhu et al. 2018; Hoffman et al. 2000; Tang, et al. 2020). LUSC often shows keratinization, pearl formation, and intercellular bridges, whereas LUAD may exhibit lepidic, glandular, papillary or micropapillary, or solid architecture (Bashir et al. 2019). These two histological subtypes always present different anatomical sites and glucose metabolism levels, and reflect the need for different optimal treatments to improve clinical outcomes (Herbst et al. 2008; Ma et al. 2018a; Zhu et al. 2018; Hoffman et al. 2000). Therefore, accurately predicting LUSC and LUAD is of paramount importance prior to clinical interventions (Ma et al. 2018a).

The first-line reference in preoperatively diagnosing LUSC and LUAD is lung biopsy (Herbst et al. 2008; Ma et al. 2018a; Zhu et al. 2018; Hoffman et al. 2000; Mahon et al. 2019), which is an invasive diagnostic approach with a high level of risks in clinical practice (Ebrahimi et al. 2016). In addition, concerning the issue of tumor heterogeneity of NSCLC, lung biopsy examines only very limited proportions of the tumor tissue and is incapable of completely characterizing tumor properties (Su 2019; Zhu et al. 2018).

Developing a noninvasive strategy for the accurate prediction of LUSC and LUAD preoperatively is desirable.

Non-invasive imaging technologies, such as computed tomography (CT) and multiparametric magnetic resonance imaging (mpMRI), have recently been widely used for the pretherapy diagnosis of NSCLC (Su 2019; Zhu et al. 2018; Sun et al. 2018; Starkov et al. 2018; Sollini et al. 2017; Shen et al. 2017). Compared with mpMRI, CT offers considerably better imaging efficiency, higher resolution, and fewer motion artifacts caused by breathing and is thus recommended in the guidelines for NSCLC screening and diagnosis (Bashir et al. 2019; Starkov et al. 2018). However, it is very challenging for clinicians to visually predict the histological subtype of NSCLC directly from CT images to discriminate between LUSC and LUAD.

In recent years, radiomics strategies have been used for the prediction of LUSC and LUAD. In 2016, Wu et al. explored a CT-based radiomics strategy with 440 features extracted, and the Naïve Baye’s classifier was used and achieved fair performance for the differentiation of LUSC and LUAD with an area under the curve (AUC) of the receiver-operating characteristic (ROC) curve of 0.72 (Wu, et al. 2016). Bashir et al. extracted 115 radiomics features from CT data and developed a prediction model based on the optimal features and random forest (RF) classifier, achieving an AUC of 0.82 for discriminating between LUSC and LUAD (Bashir et al. 2019). Chaunzwa et al. introduced the convolutional neural network (CNN) to the prediction task and developed a prediction model based on the Visual Geometry Group-16 (VGG-16) network (Chaunzwa, et al. 2018), obtaining an optimal AUC of 0.751.

In addition, some recent studies also integrated the radiomics strategy with positron emission tomography computed tomography (PET-CT) images, achieving favorable diagnostic performance in the differentiation of these two subtypes of NSCLC (Ma et al. 2018b; Koyasu et al. 2020; Ren 2020). For instance, Koyasu et al. proposed a PET-CT-based radiomics strategy with an extreme gradient boosting (XGBoost) classifier for the prediction task (Koyasu et al. 2020), achieving good performance with an AUC of 0.843.

Although these previous studies have repeatedly demonstrated the feasibility of the radiomics strategy based on CT or PET-CT for the prediction of histological subtypes of NSCLC, all the features they extracted were from the intratumoral region of the image. We are not aware of any work that has attempted to evaluate the peritumoral area outside the tumor to distinguish LUSC from LUAD. According to a recent study (Beig et al. 2019), perinodular region-based radiomics features on lung CT images effectively reflect the difference between LUAD and granulomas and accurately distinguish these two types of lung nodules. Whether the radiomics features extracted from the peritumoral region of NSCLC can reflect the significant difference between
LUSC and LUAD and further be used for the prediction task remains an open question to date.

Therefore, the first aim of this study was to investigate whether the radiomics features extracted from the peritumoral region of NSCLC could significantly reflect the difference between LUSC and LUAD. To achieve this goal, seven feature categories were employed in this study, including morphological features, histogram-based features (first-order features, hereafter), Haralick features of co-occurrence matrices (CM features, hereafter) (Haralick et al. 1973), and features derived from the run length matrix (RLM features, hereafter) (Galloway 1975), the neighborhood gray-tone difference matrix (NGTDM features, hereafter) (Amadasun and King 1989), the gray-level size zone matrix (GLSZM features, hereafter) (Thibault et al. 2014), and gray-level dependence matrix (GLDM features, hereafter) (Sun and Wee 1983) to fully characterize the global, local, and regional differences of the tissue in the peritumoral region between LUSC and LUAD (Xu, et al. 2019).

The second aim was to develop an accurate and consistent model for predicting LUSC and LUAD. To fulfill this aim, both intra- and peritumoral region-based radiomics features were utilized, and an ensemble classifier that combined multiple binary classifiers, such as support vector machine (SVM), RF, and XGBoost, was used to form a more robust predictive model. The diagnostic performance of the model was then assessed with AUC for the differentiation of LUSC and LUAD. Besides, the performance of the proposed model was also compared with that of the widely used deep network Visual Geometry Group-16 (VGG-16) (Li 2019).

Materials and methods

This retrospective study was approved by the institutional ethics review board of Xijing Hospital, and informed consent was waived. The overall methodological pipeline of this study is shown in Fig. 1.

Patients

A total of 146 archival patients with postoperatively confirmed NSCLC were collected from Xijing Hospital. The inclusion criteria were as follows: (i) primary LUSC or LUAD was pathologically confirmed; (ii) CT scan was performed prior to any therapies. Patients who met one of the following conditions were excluded: i) lack of postoperative pathological information to confirm the histopathological subtype of the patient as LUSC or LUAD (n = 21); (ii) missing preoperative CT scan (n = 16); or (iii) poor imaging quality makes accurate tumor annotations extremely difficult (n = 4). Finally, 105 subjects were eligible for this study.
study, including 47 patients with LUSC and 58 patients with LUAD. The patients were then randomly allocated into the training cohort \((n = 73)\) and testing cohort \((n = 32)\). The inclusion–exclusion process is illustrated in Fig. 2.

**Image acquisition and region of interest annotation**

All patients underwent thoracic CT imaging using a uCT 760 system (United Imaging Healthcare, China). The primary scanning parameters were as follows: 80 kV; 80 mAs; detector collimation: \(64 \times 0.6\) mm; rotation time: 0.4 s; slice thickness: 5 mm; spacing between slices: 5 mm; pixel spacing: \(0.6 \times 0.6\) mm; and matrix size, \(512 \times 512\). The entire lung region was scanned in each patient, and the image slice varied from 100 to 400.

Two types of regions of interest (ROIs), including intra- and peritumoral regions, were annotated from the CT images, as shown in Fig. 3. Prior to the intratumoral region annotation of each CT dataset, the axial image slice was selected to obtain the largest area of the archived tumor with the maximal size in each patient’s lung region. Then, a manually depicted polygonal ROI was used to segment the intratumor region on the selected image slice. Two radiologists with 20 and 10 years of lung CT interpretation experience independently performed intratumoral region delineation using a custom-developed package. Then, divergence of their delineation results was carefully corrected by consensus.

After the intratumoral region mask was obtained, we adopted the morphological dilation operator to generate a new region mask that was approximately 10 mm larger in
radial distance than the intratumoral region according to pixel size (Beig et al. 2019). Then, the corresponding peritumoral region was the ring of the lung parenchyma around the tumor that was obtained by subtracting the intratumoral region mask from the new region mask after morphological expansion, as shown in Fig. 3. Finally, the peritumoral region was further divided into two rings including the first ring (0–5 mm) and the second ring (5–10 mm) for feature extraction and comparison (Beig et al. 2019).

Radiomics feature extraction

After intra- and peritumoral ROI segmentation, ten filters, including wavelet-HL, wavelet-LL, wavelet-LH, wavelet-HH, square, square root, logarithm, exponential, gradient, and local binary pattern (LBP), were utilized to the original image to magnify the tissue patterns and unearth important features. Then, six feature categories, including first-order features, GLCM features, GLRLM features, NGTDM features, GLSZM features, and GLDM, were calculated from the original segmented image data and ten generated images of the intratumoral and two rings of the peritumoral regions (Zwanenburg, et al. 2020). Given that the peritumoral region was dilated based on use of the intratumoral region, the shape 2D features were only calculated from the intratumoral region. Therefore, 1032, 1023, and 1023 radiomics features were extracted from the intratumoral region and the first ring and the second ring of the peritumoral region, respectively, as shown in Table 1. Open source Pyradiomics (version 3.0.1) was used to perform this analysis (Griethuysen et al. 2017). All of the codes and results have been attached in the Appendix document in supplementary material.

Feature selection

In this study, a two-step feature selection strategy was adopted to determine an optimal subset of features for model construction, as shown in Fig. 1. The first step was statistical analysis of all these features between LUSC and LUAD, which was performed with Scikit-learn. Student’s $t$ test with a significant $p$-value set as 0.05 was then performed with all radiomics features to select those with significant intergroup differences between LUSC and LUAD (Probable et al. 1992).

Then, all significant features were standardized to eradicate differences of the feature-value scales. The normalized feature $z$ of each feature $x$ for a specific patient is calculated as follows:

$$z = \frac{x - \overline{x}}{\sigma},$$  

(1)

where $\overline{x}$ and $\sigma$ are the mean and standard deviation, respectively, of each feature from the training cohort.

In the second step of feature selection, three widely used feature selection algorithms, including the minimum redundancy maximum relevance method (mRMR) (Peng et al. 2005), the least absolute shrinkage and selection operator(LASSO) (Tibshirani 1996; Sauerbrei et al. 2007), and the linear SVM-based recursive feature elimination

| Characteristics                          | Training cohort ($n = 73$) | Testing cohort ($n = 32$) | $P$ value |
|------------------------------------------|----------------------------|---------------------------|-----------|
| Age, years                               | 61 [35, 76]                | 59 [39, 83]               | 0.87a     |
| Median (range)                           |                            |                           |           |
| Sex, no. (%)                             | 54 / 73 (73.97%)           | 24 / 32 (75.00%)          | 0.91b     |
| Male                                     |                            |                           |           |
| Female                                   | 19 / 73 (26.03%)           | 8 / 32 (25.00%)           |           |
| Smoking, no. (%)                         | 49 / 73 (67.12%)           | 20 / 32 (62.50%)          | 0.65b     |
| Yes                                      |                            |                           |           |
| No                                       | 24 / 73 (32.88%)           | 12 / 32 (37.50%)          |           |
| Side, no. (%)                            | 22 / 73 (30.14%)           | 10 / 32 (31.25%)          | 0.90b     |
| Upper left lobe                          |                            |                           |           |
| Lower left lobe                          | 12 / 73 (16.44%)           | 4 / 32 (12.50%)           |           |
| Upper right lobe                         | 20 / 73 (27.40%)           | 7 / 32 (21.88%)           |           |
| Middle right lobe                        | 2 / 73 (2.74%)             | 1 / 32 (3.13%)            |           |
| Lower right lobe                         | 17 / 73 (23.29%)           | 10 / 32 (31.25%)          |           |
| Histopathological subtype, no. (%)       | 33 / 73(45.21%)            | 14 / 32(43.75%)           | 0.89b     |
| Squamous cell carcinoma (LUSC)           |                            |                           |           |
| Adenocarcinoma (LUAD)                    | 40 / 73(54.79%)            | 18 /32(56.25%)            |           |

*a Student's $t$ test  
*b Chi-square test
(SVM-RFE) (Fehr et al. 2015), were further implemented with these significant features to select an optimal feature subset from the training cohort for model development and external testing.

**Model development based on ensemble learning and validation**

With optimal features selected, the predictive model was then developed using the training cohort and the ensemble learning strategy with tenfold cross-validation and 10 rounds, as shown in Fig. S1 of the Appendix. And the performance of the model was then externally evaluated using the testing cohort. In each split, ninefold were used for model training and the fold remained was used for performance validation. The training performance we finally obtained was the average value of all the validation with ten splits. Then, we repeated the entire process with ten rounds to obtain optimal hyperparameters with the best average performance. After that, the entire training cohort was used for model development with these optimal hyperparameters. And the testing cohort, which was not participating in the training process, was used as the external cohort to verify the overall performance.

Five commonly used binary classifiers, including the quadratic discriminant analysis (QDA) classifier, SVM with radial basis function (RBF) kernel, SVM with sigmoid/tanh kernel, RF, and XGBoost, were included in the ensemble learning framework. QDA is the most commonly used binary classifier, which has no same-covariance assumption for each binary class (Linear and Quadratic Discriminant analysis 2022; Tharwat 2016). SVM is a classical machine learning classifier with several typical kernels, such as RBF and sigmoid/tanh, that is used to compute the decision boundary that separates two classes with the maximum marginal distance (Hastie et al. 2009; Lam, et al. 2012; Stenzinger, et al. 2021). It has advantages in dealing with nonlinear features and is not easily overfit with even small datasets (Liang et al. 2018). The RF classifier can build multiple random decision trees (100 trees of the default parameter in Scikit-learn to avoid overfitting) and integrate them to make an accurate diagnosis (Liang et al. 2018; Khalilia et al. 2011; Seera and Lim 2014). XGBoost offers many benefits in classification, including high precision and consistency and the prevention of overfitting (Chen and Guestrin 2016; Colen 2021); thus, it was also included in the ensemble learning strategy.

The ensemble classifier was finally developed by weighting the predictive value of these five classifiers in the model training process, which can be expressed as follows:

\[
P(j) = \sum_{i=1}^{5} \omega_i p_i(j), \tag{2}
\]

where \(P(j)\) represents the final predictive value of the \(j\)th patient; \(p_i(j)\) denotes the predictive value of the \(j\)th patient using the \(i\)th classifier; and \(\omega_i\) is the weighting parameter of the \(i\)th classifier in the ensemble learning process, which meets the following condition:

\[
\sum_{i=1}^{5} \omega_i = 1. \tag{3}
\]

In this study, the optimal weight \(\omega_i\) was determined based on minimizing the predictive error in the training process, and the cutoff \(P(j)\) for assigning the patient to the LUAD group was set as 0.5. If \(P(j)\) was greater than or equal to 0.5, the \(j\)th patient was allocated to the LUAD group. The overall performance was evaluated using both the training cohort and the testing cohort with the quantitative metrics of accuracy and AUC (Gupta and Mittal 2019a, b, c; Kora and Krishna 2014).

In addition, we also compared the performance of the proposed ensemble classifier with the VGG-16 network. The experiment was conducted with an *NVIDIA GeForce RTX 3090* machine with 24 GB of memory. Hyperparameter parameters included: epoch of 50, batch size of 8, and learning rate of 0.0001. The optimizer is the Adam optimizer.

**Statistical analysis**

Statistical analyses of the patient demographics were performed using IBM SPSS statistics (version 19.0, Armonk, NY), and Python software (version 3.6 DL-GPU) was used to perform statistical selection of features with significant differences between LUSC and LUAD. Chi-square tests were performed to evaluate significant differences in primary clinical factors distributed between the training and testing cohorts, and Student’s \(t\) tests were used to select significant radiomics features between LUSC and LUAD. Two-sided \(p\) values less than 0.05 were considered significant (Xu, et al. 2019; Wu et al. 2017; Wu et al. 2018).

**Results**

**Demographics of eligible patients**

A total of 105 NSCLC patients were eligible for this study, including 47 patients with LUSC and 58 with LUAD. These patients were randomly allocated into the training cohort \((n = 73)\) and the testing cohort \((n = 32)\). The baseline demographics and clinical information of these patients were collected from the archival medical document, as shown in Table 1. Statistical analyses indicate no significant differences between both the training and testing cohorts in terms of all these primary factors.
Fig. 4 Statistical analysis-based feature selection results: a all 839 significant features from intra- and peritumoral regions; b 500 significant features from the intratumoral region; c 220 significant features from the first ring (0–5 mm) of peritumoral region; d 119 significant features from the second ring (5–10 mm) of peritumoral region.

Fig. 5 Intra- and peritumoral tissue distribution differences between LUSC and LUAD characterized by the significant radiomics feature energy on CT images with the unit normalized as “1” on the color bar.
Results of the two-step feature selection strategy

A total of 3078 standardized radiomics features, including 1032 features from the intratumoral region, 1023 from the first ring (0–5 mm), and 1023 from the second ring (5–10 mm) of peritumoral regions, were analyzed using Student’s t-test (p value < 0.05) to determine those with significant intergroup differences between LUSC and LUAD. Eventually, 500 significant features were selected from the intratumoral region, whereas only 220 and 119 significant features were selected from the first ring and second ring of peritumoral regions, respectively, as shown in Fig. 4. These results indicate that (i) a large number of radiomics features extracted from the peritumoral region can also reflect the significant differences in tissue distribution patterns between LUSC and LUAD; (ii) the closer the peritumoral region is located to the intratumoral region, the more features with significant differences could be obtained to reflect the tumor property difference. Figure 5 illustrates an example of the intra- and peritumoral tissue distribution differences of LUSC and LUAD determined using one of the significant radiomics features, energy, with 3 × 3 sliding patches on the CT image.

After statistical analysis-based feature selection, three radiomics feature subsets were finally obtained, including (i) 500 significant features from the intratumoral region, (ii) 339 significant features from the entire peritumoral region, and (iii) 839 significant features from both intratumoral and peritumoral regions. All these significant features in each feature subset were further selected using three commonly applied strategies: SVM-RFE, LASSO, and mRMR with the mutual information difference (MID), as shown in Figs. 6, 7, 8. Table 2 shows the results after the second-step feature selection procedure.

Classification model development and performance evaluation

As these optimal feature subsets were determined, classification models were developed using five commonly used machine learning classifiers and the ensemble classifier with the training cohort, and the performance of each model was evaluated using both training and testing cohorts for distinguishing LUSC from LUAD. The results are presented in Fig. 9. Three columns of subfigures in Fig. 9 exhibit the performance of predictive models developed using
optimal feature subsets determined from the intratumoral region, peritumoral region, and both intra- and peritumoral regions. These findings indicate that (i) the classification model determined from the peritumoral region achieved comparable performance to that from the intratumoral region; (ii) the classification model determined from intra- and peritumoral regions dramatically improved the overall performance for the prediction of LUSC and LUAD; and
which is obvious inferior to that of our proposed model. Testing cohort for the identification of LUSC and LUAD, proposed model. Finally, it obtained an AUC of 0.67 in the dated with the same testing cohort as those were used in our model was trained with the same training cohort and validation model that was developed using the ensemble classifier for classification model building. The results of development. Five independent classifiers, QDA, SVM with RBF kernel, SVM with sigmoid/tanh kernel, RF, and XGBoost, which were reported to have favorable classification performance and robustness for the diagnosis of cancer phenotypes with a small database, were utilized to form an ensemble classifier for classification model building. The results of the model that was developed using the ensemble classifier and optimal features selected by SVM-RFE from intra- and peritumoral regions demonstrate favorable discriminative power with both the training and testing cohorts.

In recent years, CT-/PET-CT/multimodal MRI-based radiomics strategies have been repeatedly demonstrated to have great capability for the prediction of LUSC and LUAD (Bashir et al. 2019; Tang et al. 2020; Wu et al. 2016; Ma et al. 2018b; Koyasu et al. 2020; Ren 2020). The diagnostic performance ranged between 0.72 and 0.843. Nevertheless, all these previous studies only focused on how to extract an increasing number of features from the intratumoral region of the image, regardless of the peritumoral parenchyma, which might also contain substantial information and be of equal importance for the prediction task. Some studies have revealed that the interface of the tumor has a “rim” of densely packed tumor-infiltrating lymphocytes and tumor-associated macrophages in representative hematoxylin and eosin-stained images (Hoffman et al. 2000; Beig et al. 2019; Kirienko et al. 2018; Jong et al. 2018). At a macroscopic scale, the densely packed stromal tumor-infiltrating lymphocytes around LUAD represent fine and smooth textures on CT images and thus could be potential imaging biomarkers for the identification of LUAD from LUSC (Beig et al. 2019). However, whether radiomics features extracted from the peritumoral parenchyma region effectively reflect the intergroup difference of the tissue and microenvironment between LUSC and LUAD remains unknown to date.

In this study, we investigated the feasibility of CT-based radiomic features extracted from intra- and peritumoral regions of NSCLC to reflect the tissue distribution differences between LUSC and LUAD, and developed a CT-based radiomics strategy that incorporated high-throughput features with an ensemble classifier for the preoperative prediction of LUSC and LUAD. Three widely used methods, SVM-RFE, LASSO, and mRMR, were employed to select optimal features with significant intergroup differences between LUSC and LUAD for classification model development. Five independent classifiers, QDA, SVM with RBF kernel, SVM with sigmoid/tanh kernel, RF, and XGBoost, which were reported to have favorable classification performance and robustness for the diagnosis of cancer phenotypes with a small database, were utilized to form an ensemble classifier for classification model building. The results of the model that was developed using the ensemble classifier and optimal features selected by SVM-RFE from intra- and peritumoral regions demonstrate favorable discriminative power with both the training and testing cohorts.

(iii) the model developed by the ensemble classifier achieved more favorable and consistent performance with training and testing cohorts compared with those developed by five independent classifiers. Table 3 shows the performance of classification models developed by the ensemble classifier for the prediction task, indicating that the ensemble classification model developed by SVM-RFE-based optimal features determined from intra- and peritumoral regions achieved the best performance with AUC values of 0.87 and 0.78 in the training and testing cohorts, respectively. Besides, the VGG-16 model was trained with the same training cohort and validated with the same testing cohort as those were used in our proposed model. Finally, it obtained an AUC of 0.67 in the testing cohort for the identification of LUSC and LUAD, which is obvious inferior to that of our proposed model.

Discussion

In this study, we investigated the feasibility of CT-based radiomic features extracted from intra- and peritumoral regions of NSCLC to reflect the tissue distribution differences between LUSC and LUAD, and developed a CT-based radiomics strategy that incorporated high-throughput features with an ensemble classifier for the preoperative prediction of LUSC and LUAD. Three widely used methods, SVM-RFE, LASSO, and mRMR, were employed to select optimal features with significant intergroup differences between LUSC and LUAD for classification model development. Five independent classifiers, QDA, SVM with RBF kernel, SVM with sigmoid/tanh kernel, RF, and XGBoost, which were reported to have favorable classification performance and robustness for the diagnosis of cancer phenotypes with a small database, were utilized to form an ensemble classifier for classification model building. The results of the model that was developed using the ensemble classifier and optimal features selected by SVM-RFE from intra- and peritumoral regions demonstrate favorable discriminative power with both the training and testing cohorts.

In recent years, CT-/PET-CT/multimodal MRI-based radiomics strategies have been repeatedly demonstrated to have great capability for the prediction of LUSC and LUAD (Bashir et al. 2019; Tang et al. 2020; Wu et al. 2016; Ma et al. 2018b; Koyasu et al. 2020; Ren 2020). The diagnostic performance ranged between 0.72 and 0.843. Nevertheless, all these previous studies only focused on how to extract an increasing number of features from the intratumoral region of the image, regardless of the peritumoral parenchyma, which might also contain substantial information and be of equal importance for the prediction task. Some studies have revealed that the interface of the tumor has a “rim” of densely packed tumor-infiltrating lymphocytes and tumor-associated macrophages in representative hematoxylin and eosin-stained images (Hoffman et al. 2000; Beig et al. 2019; Kirienko et al. 2018; Jong et al. 2018). At a macroscopic scale, the densely packed stromal tumor-infiltrating lymphocytes around LUAD represent fine and smooth textures on CT images and thus could be potential imaging biomarkers for the identification of LUAD from LUSC (Beig et al. 2019). However, whether radiomics features extracted from the peritumoral parenchyma region effectively reflect the intergroup difference of the tissue and microenvironment between LUSC and LUAD remains unknown to date.

In this study, we found that a large number of radiomics features extracted from the intratumoral region and peritumoral region were significantly different between LUSC and LUAD, and the total number of significant features extracted from the first ring (0–5 mm) peritumoral region was much greater than that of the significant features extracted from the second ring (5–10 mm) peritumoral region. These results demonstrate and verify for the first time the hypothesis that the peritumoral region on CT images also contains substantial information that can reflect the tissue texture difference between LUSC and LUAD. In addition, the closer the peritumoral region is to the intratumoral region, the more substantial the information it contains.

Most of the previous studies only focused on extracting features from the original image data, neglecting the image filters that not only reduce the noise but also enhance the quality and magnify the texture in the image (Xu et al. 2017a, b). Therefore, in this study, ten filters, including wavelet-HL, wavelet-LL, wavelet-LH, wavelet-HH, square, square root, logarithm, exponential, gradient, and LBP, were utilized to preprocess the image for feature extraction. Seven categories of radiomics features, including morphological features, first-order features, second-order features, and high-order texture features, were adopted in this study to fully characterize the shape properties and global, local, and regional distribution patterns of the tissue, respectively. Student’s t tests integrated with three widely applied feature selection algorithms (SVM-RFE, LASSO, and mRMR), were adopted for optimal feature selection and performance comparison. The results indicate that the optimal features selected using the SVM-RFE algorithm from all significant features of both intra- and peritumoral regions have the most powerful diagnostic ability for the discrimination between LUSC and LUAD.
Classification model development is the last but most crucial step in the proposed radiomics strategy for the prediction of LUSC and LUAD. In this step, the choice of an optimal decision classifier, for instance, SVM with RBF kernel or Sigmoid kernel, RF, QDA, or XGBoost represent the core influence of performance variation (Liang et al. 2018). Hence, the determination of an optimal classifier is of critical importance. To fully integrate all the merits of these five independent classifiers, an ensemble classifier was generated using five independent classifiers, SVM with RBF kernel, or sigmoid kernel, RF, QDA, and XGBoost, and its diagnostic performance was compared with these independent classifiers. The results indicate that (i) the classification model developed using the ensemble classifier achieves the most favorable, consistent and robust diagnostic performance compared with other independent classifiers; and (ii) optimal features determined by SVM-RFE from both intra- and peritumoral regions with the ensemble classifier achieve the best diagnostic performance for the prediction of LUSC and LUAD with both training and testing cohorts. In addition, the classification results of all these models developed by each classifier with optimal features determined from intra-tumoral, peritumoral, or both of intratumoral and peritumoral regions using SVM-RFE, LASSO, and mRMR also revealed that although the model based on the ensemble classifier did not always obtain the best results, it always ranked as one of the top two models in terms of the AUC with both cohorts, suggesting remarkable consistency and robustness in the prediction of LUSC and LUAD.

We further included the performance comparison between our methodology with the deep-learning algorithm VGG-16 which has been widely used for NSCLC image analysis. VGG-16 uses more channels, deeper convolutional layers, and wider feature map, which can extract more representative features for disease characterization. In addition, small kernels (3 x 3) were utilized in VGG-16 to replace the large kernels in other deep network, which can largely reduce the amount of parameters, carry out more nonlinear mapping, and help to increase the fitting ability of the network. Results indicate that the propose model has more advantages in the diagnosis of LUSC and LUAD with very limited data size.

However, the results of this study should be carefully interpreted due to the following limitations. First of all, the sample size of our study is small and single-centered, which might impair the generalizability of the model for the large multi-center database application. Moreover, other potential clinical factors, such as gene mutations and key molecular biomarkers, were not included in the current study given the incomplete data in the archival database, which should be further analyzed. In addition, deep radiomics features incorporating the current manual radiomics features might further improve current performance in the prediction of LUSC and LUAD. In future work, a large database from multiple centers will be collected for further evaluating the proposed method. Besides, multimodal imaging data like PET/CT or PET/MR will be considered to further improve the diagnostic performance.

In conclusion, the proposed CT-based radiomics strategy that extracts features from intra- and peritumoral regions, adopts SVM-RFE for optimal feature selection, and utilizes ensemble learning for classification model development is demonstrated with favorable predictive precision and stability for preoperatively prediction of LUSC and LUAD.

### Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s00432-022-04015-z.

## Author contributions

XX, XT, and HH contributed to the study concept, design, and data interpretation. XT contributed to the CT and clinical data collection. XT and HY contributed to the intratumoral region annotation. HH, XX and PD performed the peritumoral region extraction and radiomics feature calculation; XX, HH and XT contributed to the model construction and data analysis. XX, XT, and HH contributed to the manuscript drafting, editing and revision. All authors approve the final version of the manuscript for submission.

### Funding

This work was funded by the National Natural Science Foundation of China (No. 81901698) and Young Eagle plan of High Ambition Project (No. 2020CYJHXXP).

### Data availability statement

The raw/processed data of this study cannot be publicly shared at present as it forms part of an ongoing study, but it could be available under reasonable request from the corresponding author with the permission of the Institutional Review Board. Results and code package in each step of this study have been arranged in a document named as “Appendix”. The code package has also been uploaded to Gitee for publicly sharing and further perfection (https://gitee.com/ yang-tianran-01/radiomics-ensemble_learning/commit/ d51e6859ef48e92c0c794639f08286ace89569f8).

| Method            | Classifier | From the intratumoral region | From the peritumoral region | From intra- and peritumoral regions |
|-------------------|------------|------------------------------|-----------------------------|-------------------------------------|
|                   |            | Training | Testing | Training | Testing | Training | Testing |
| SVM-RFE           | Ensemble   | 0.87     | 0.66    | 0.83     | 0.60    | 0.87     | 0.78    |
| LASSO             | Ensemble   | 0.76     | 0.63    | 0.73     | 0.64    | 0.79     | 0.68    |
| mRMR with MID     | Ensemble   | 0.77     | 0.68    | 0.64     | 0.56    | 0.73     | 0.71    |

*The AUC value with bold ranks as the top place in each column*
Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This study was approved by the institutional ethics review board of Xijing Hospital, and informed consent was waived.

References

Amadasun M, King R (1989) Textural features corresponding to texural properties. IEEE Trans Syst Man Cybern 19(5):1264–1274

Bashir U et al (2019) Non-invasive classification of non-small cell lung cancer: a comparison between random forest models utilising radiomic and semantic features. Br J Radiol 92(20190159):1–8

Beig N et al (2019) Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas. Radiology 290(3):783–792

Bray F et al (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin. https://doi.org/10.3322/caac.21492

Chaunzwa TL et al (2018) Using deep-learning radiomics to predict lung cancer histology. J Clin Oncol 36(15_Suppl):8545–8545

Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: 2016. Association for Computing Machinery

Colen RR et al (2021) Radiomics analysis for predicting pembrolizumab response in patients with advanced rare cancers. J Immunother Cancer 9(4):1752

de Jong EEC et al (2018) Applicability of a prognostic CT-based radiomic signature model trained on stage I-III non-small cell lung cancer in stage IV non-small cell lung cancer. Lung Cancer 124:6–11

Ebrahimi M et al (2016) Diagnostic concordance of non–small cell lung carcinoma subtypes between biopsy and cytokine specimens obtained during the same procedure. Cancer Cytopathol 124(10):737–743

Fehr D et al (2015) Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. Proc Natl Acad Sci 112(46):E6265–E6273

Galloway MM (1975) Texture analysis using gray level run lengths. Comput Graph Image Process 4:172–179

Gupta V, Mittal M (2019a) R–peak detection in ECG signal using yule-walker and principal component analysis. IETE J Res 67:921–934

Gupta V, Mittal M (2019b) A comparison of ECG signal pre-processing using FrFT, FrWT and IPCA for improved analysis. IRBM 40:145–156

Gupta V, Mittal M (2019c) QRS complex detection using STFT, chaos analysis, and PCA in standard and real-time ECG databases. J Inst Eng (India) 100:489–497

Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. IEEE Trans Sys Man Cybern 3(6):610–621

Hastie T, Tibshirani R, Friedmann J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer Science & Business Media, Berlin, p 757

Herbst RS, Heymach JV, Lippman SM (2008) Lung cancer. N Engl J Med 359(13):1367–1380

Hoffman PC, Mauer AM, Vokes EE (2000) Lung cancer. Lancet 355(9202):479–485

Khailila M, Chakraborty S, Popescu M (2011) Predicting disease risks from highly imbalanced data using random forest. BMC Med Inform Decis Mak 11(1):51

Kirienko M et al (2018) Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery. Eur J Nucl Med Mol Imaging 45(2):207–217

Kora P, Krishna KSR (2014) Myocardial infarction detection using magnitude squared coherence and support vector machine. Med Imaging. https://doi.org/10.1109/MedCom.2014.7006037

Koyasu S et al (2020) Usefulness of gradient tree boosting for predicting histological subtype and EGFR mutation status of non-small cell lung cancer on 18F FDG-PET/CT. Ann Nucl Med 34(1):49–57

Lam H-K et al (2012) Computational intelligence and its applications: evolutionary computation, fuzzy logic, neural network and support vector machine techniques. World Scientific, London, p 318

Li S et al (2019) Predicting lung nodule malignancies by combining deep convolutional neural network and handcrafted features. Phys Med Biol 64(17):175012

Liang C et al (2018) A computer-aided diagnosis scheme of breast lesion classification using GLGLM and shape features: Combined-view and multi-classifiers. Phys Med 55:61–72

Linear & Quadratic Discriminant Analysis · UC Business Analytics R Programming Guide. https://uc-r.github.io/discriminant_analysis

Ma Y et al (2018a) Intra-tumoural heterogeneity characterization through texture and colour analysis for differentiation of non-small cell lung carcinoma subtypes. Phys Med Biol 63(16):1658

Ma Y et al (2018b) Intra-tumoural heterogeneity characterization through texture and colour analysis for differentiation of non-small cell lung carcinoma subtypes. Phys Med Biol 63(16):1658

Mahon RN, Hugo GD, Weiss E (2019) Repeatability of texture features derived from magnetic resonance and computed tomography imaging and use in predictive models for non-small cell lung cancer outcome. Phys Med Biol 64:145007

Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27(8):1226–1238

Probable T (1992) Error of a mean. In: Kotz S, Johnson NL (eds) Breakthroughs in statistics: methodology and distribution. Springer, New York, pp 33–57

Ren C et al (2020) Machine learning based on clinico-biological features integrated 18F-FDG PET/CT radiomics for distinguishing squamous cell carcinoma from adenocarcinoma of lung. Eur J Nucl Med Mol Imaging. https://doi.org/10.1007/s00259-020-05065-6

Sauerbrei W, Royston P, Binder H (2007) Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Stat Med 26(30):5512–5528

Seera M, Lim CP (2014) A hybrid intelligent system for medical data classification. Expert Syst Appl 41(5):2239–2249

Shen C et al (2017) 2D and 3D CT radiomics features prognostic performance comparison in non-small cell lung cancer. Transl Oncol 10(6):886–894

Sollini M et al (2017) PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology. Sci Rep 7(1):358

Starkov P et al (2018) The use of texture-based radiomics CT analysis to predict outcomes in early-stage non-small cell lung cancer treated with stereotactic ablative radiotherapy. Br J Radiol 91(20180228):1–7

Stenziung A et al (2021) Artificial intelligence and pathology: from principles to practice and future applications in histomorphometry and molecular profiling. Semin Cancer Biol. https://doi.org/10.1016/j.semcancer.2021.02.011

Su R et al (2019) Identification of expression signatures for non-small cell lung carcinoma subtype classification. Bioinformatics. https://doi.org/10.1093/bioinformatics/btz557
Sun C, Wee WG (1983) Neighboring gray level dependence matrix for texture classification. Compute vis Graph Image Process 23:341–352
Sun W et al (2018) Effect of machine learning methods on predicting NSCLC overall survival time based on Radiomics analysis. Radiat Oncol 13(1):197
Sung H et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin 71:209–249
Tang X et al (2020) Elaboration of a multimodal MRI-based radiomics signature for the preoperative prediction of the histological subtype in patients with non-small-cell lung cancer. BioMed Eng Online. https://doi.org/10.1186/s12938-019-0744-0
Tharwat A (2016) Linear vs quadratic discriminant analysis classifier: a tutorial. Int J Appl Pattern Recogn 3(2):145
Thibault G, Angulo J, Meyer F (2014) Advanced statistical matrices for texture characterization: application to cell classification. IEEE Trans Biomed Eng 61(3):630–637
Tibshirani R (1996) Regression Shrinkage and Selection Via the Lasso. J Roy Stat Soc Ser B (methodol) 58(1):267–288
van Griethuysen JJM et al (2017) Computational radiomics system to decode the radiographic phenotype. Can Res 77(21):e104–e107
Wu W et al (2016) Exploratory study to identify radiomics classifiers for lung cancer histology. Front Oncol. https://doi.org/10.3389/ fonc.2016.00071
Wu S et al (2017) A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. Clin Cancer Res 23(22):6904–6911
Wu S et al (2018) Development and validation of an MRI-based radiomics signature for the preoperative prediction of lymph node metastasis in bladder cancer. EBioMedicine 34:76–84
Xu X et al (2017a) Preoperative prediction of muscular invasiveness of bladder cancer with radiomic features on conventional MRI and its high-order derivative maps. Abdom Radiol (NY) 42(7):1896–1905
Xu X et al (2017b) Three-dimensional texture features from intensity and high-order derivative maps for the discrimination between bladder tumors and wall tissues via MRI. Int J Comput Assist Radiol Surg 12(4):645–656
Xu X et al (2019) A predictive nomogram for individualized recurrence stratification of bladder cancer using multiparametric MRI and clinical risk factors. J Magn Resonance Imaging 50:1893–1904
Zhu X et al (2018) Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. Eur Radiol 28(7):1–7
Zwanenburg A et al (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 295(2):328–338

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.