Efficacy of radiomics model based on the concept of gross tumor volume and clinical target volume in predicting occult lymph node metastasis in non-small cell lung cancer

Zeng Chao
Clinical Medicine College, North China University of Science and Technology

Wei Zhang
The Affiliated Hospital of Qingdao University

Meiyue Liu
Clinical Medicine College, North China University of Science and Technology

Jianping Liu
Tangshan People's Hospital

Qiangxin Zheng
Clinical Medicine College, North China University of Science and Technology

Jianing Li
Clinical Medicine College, North China University of Science and Technology

Zhiwu Wang (✉ tcm2000@163.com)
Tangshan People's Hospital

Guogui Sun
Clinical Medicine College, North China University of Science and Technology

Article

Keywords: Radiomics, Occult lymph node metastasis, Non-small cell lung cancer, Chest-enhanced CT, Prediction model

Posted Date: December 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2346906/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Objective: To extract the gross tumor volume (GTV) and clinical target volume (CTV) of lung cancer from computed tomography (CT) images by radiomics method. To establish a predictive model for occult Lymph node metastasis (LNM) in patients with stage - A non-small cell lung cancer (NSCLC) based on contrast-enhanced CT.

Methods: A retrospective analysis of 598 patients in different hospitals with preoperative enhanced CT presentation of stage I-IIA NSCLC was performed. Of all patients, 119 patients (20%) had postoperative pathology suggestive of LNM. On chest-enhanced CT (CECT) arterial phase pictures, GTV and CTV radiomics characteristics were retrieved. The training group included 401 patients (67 with LNM and 334 without LNM), and the external validation group included 197 patients (52 with LNM and 145 without LNM). For choosing radiomics features, the LASSO regression algorithm was applied. Three models (GTV model, CTV model, and GTV+CTV hybrid model) were developed to predict occult LNM in early NSCLC patients using the screened features. The models were implemented by plotting nomogram plots.

Results: Eight optimal radiomics features were finally locked and significantly correlated with occult LNM (all P<0.05). The ROC curves of the three models showed good predictive effects. The AUC values of the GTV and CTV models in the validation group were 0.821 and 0.812. The AUC of the GTV+CTV hybrid model was significantly better than that of the CTV model (the AUC of the training group and the validation group in the GTV+CTV hybrid model was 0.869 and 0.906, respectively). Moreover, the GTV model (the AUC of the training group and the validation group in the GTV model were 0.845 and 0.821), and statistically significant by Delong test. ROC curve and decision curve showed that the radiomics model had clinical application value.

Conclusions: The radiomics prediction model based on GTV and CTV was developed and verified in this study to be able to predict occult LNM in patients with preoperative clinical stage I-IIA NSCLC, and it was found that the combined model is significantly more accurate than either model operating independently.

1. Introduction

Lung cancer has been identified as the leading cause of cancer death worldwide, accounting for 18.0% of all cancer deaths and the leading cause of death in men. Non-small cell lung cancer accounts for 80%-85% of all lung cancer cases [1]. Low-dose computed tomography (CT) screening programs can identify the majority of patients with early-stage non-small cell lung cancer (NSCLC). Surgery combined with systemic therapy is the primary treatment option for these patients with early resectable NSCLC [2,3]. Currently, anatomical lobectomy and lung mediastinal lymph node cleaning are the recommended treatments for non-small cell lung cancer. Early lung lobectomy and selective lymph node cleaning treatment effects are also significant, while a larger range of lymph node cleaning can reduce the risk of cancer lymph node metastasis, however, physical trauma to the patient also greatly increased. Such as unnecessary normal tissue damage, pneumothorax, lymph node leakage, prolonged hospital stay and other risks [4]. With the popularization of low-dose CT screening, the proportion of early lung cancer screened has been increasing [5]. Some studies have reported that the incidence of LNM in early lung cancer cases is about 10% [6], and some cases are occult lymph node metastases (LNM) cases. Consequently, if preoperative imaging can reliably identify negative lymph nodes, many patients can avoid lymph node dissection.
surgery, lower surgical risks, and include stereotactic radiotherapy and other local treatment options in the
treatment of early lung cancer. At present, CT, magnetic resonance imaging (MRI), and positron emission
computed tomography (PET-CT) are commonly used clinically for the non-invasive evaluation of lymph nodes,
among which PET-CT is the best method to evaluate lymph node status. However, if the patient is complicated
with pneumonia and lymph node inflammation, the probability of a false positive of PET-CT will increase\[7-9\]. In
addition, due to the high radiation dose and high scanning cost, PET-CT is not suitable for routine preoperative
examination of NSCLC. A line of lymph node metastasis in patients with invasive lymph node dissection will
result in excessive treatment, and invasive tools like ultrasound-guided puncture and thoracoscopy, while they can
provide a more accurate assessment of lymph node status, damage in patients with normal tissue also a
significant concern. These complications include bleeding, pneumothorax, nerve damage, and increased risk of
complications like these\[10, 11\].

In recent years, the wave of precision medical radiomics received widespread attention and application\[12-
14\]. Currently, some based on clinical and tumor primary focal imaging under CECT omics characteristics build a
predictive model of lung cancer lymph node metastasis, and this model successfully achieved a good prediction
effect. Radiomics is extracted from medical images high flux data and method of analyzing the characteristic of
the different dimensions of quantitative\[15,16\]. The purpose of this study is to quantitatively describe the features
of GTV, CTV, and mixed models in CECT using radiomics analysis, and to develop and validate a radiomics
prediction model based on preoperative CECT images of stage - A NSCLC. It includes the radiomics
characteristics of the tumor itself (GTV) and clinical target volume (CTV) and the effect of combining the two
radiomics models.

2. Methods

2.1 Patients

This study was retrospective, all the data were deidentified data and no experiments were performed on the
patient. Hence the People's Hospital Institutional Review Board of Tangshan allowed waive informed consent. All
methods were carried out in accordance with relevant guidelines and regulations. All experimental protocols were
approved by the People's Hospital Institutional Review Board of Tangshan. All data is included in the
supplementary files. A total of 401 patients with lung cancer undergoing radical surgery in Tangshan People's
Hospital from September 2017 to December 2019 and 197 patients in Yantai Yuhuangding Hospital from 2016 to
2020 were collected. The age was 29-83 years old, with a median age of 62 years old, including 273 males and
325 females. Lymph node metastasis or no lymph node metastasis were classified as postoperative pathological
results. The lymph node status data of the cases in the experiment were divided into Table (1-1). Inclusion criteria:
[I] Patients with preoperative clinical staging of stage - A (UICC eighth edition TNM staging) underwent radical
surgery and systemic lymph node dissection. [II] Postoperative pathology was NSCLC with or without lymph node
metastasis. [III] All patients received Contrast-enhanced computed tomography (contrast-enhanced
computed tomography) with the same type of CT equipment one month before the operation. [IV] No signs of
lymph node metastasis (CT showed a lymph node short diameter of less than 1cm). Excluded criteria: [I] Patients
had received neoadjuvant therapy before surgery; [II] Combined with other malignant tumors; [III] Distant
metastasis.

2.2 Data Collection
**Imaging Data Acquisition:** All of the cases that were included underwent CE-CT scanning. The patient was examined while lying on his back and had his chest and diaphragm scanned. Scanning setting: Scanning thickness 5.0mm, tube voltage 120kVP, tube current 80-300mAs. All images were displayed in the standard lung (width, 1200HU; level, 600HU) and mediastinal (width, 350HU; level, 40HU) Windows settings. An iodinated contrast agent of 1.5-2ml/kg was injected through the cubital vein at a rate of 3ml/s, and the venous phase scan was delayed by the 90s. All images were saved to a hard drive in the DICOM (Digital Imaging and Medical Communication) format. Images were then imported into the Accu Contour software by Manteia (version 3.1), where GTV was manually delineated by radiologists with more than five years of experience who were not aware of any clinical or pathological information. Tumor GTV includes tumor parenchyma, interstitial blood vessels, and vacuoles within nodules, excluding normal lung tissue. The experiment's GTV-expand-1 model of the CTV, which was based on GTV, involved an outward expansion of 5 mm in six directions (Figure 1). The outer expansion part included the surrounding environment of the tumor. The manual repair was performed if the outer expansion came into contact with the chest wall, diaphragm, main trachea, great blood vessels, or heart. Data are available in supplementary materials and are available for research.

**Radiomics Feature Extraction:** The "Radiomics" tool kit in AccuContour software was used to extract the radiomics features of GTV and CTV delineated. It includes 14 Shape features and 252 First order statistics), 336 gray level co-occurrence matrix (GLCM) features, 70 neighborhood gray level difference matrix (Ngtdm) features, 180 running length matrix (RLM) features, 224 gray level region matrix (GLSZM) features and 238 gray level dependence matrix (GLDM) features, a total of 1314 radiomics features.

**2.3 Model Construction**

The cases were divided into a training group and validation group according to in-hospital and out-of-hospital with a ratio of 2:1. The model was constructed using the training group, and it was validated using the external validation group. A total of three prediction models were built by incorporating different feature types, namely, the GTV radiomics model, the CTV radiomics model, and the GTV+CTV hybrid radiomics model.

**2.4 Analysis of Data**

Statistical analysis was performed using R and RStudio statistical software (R-4.1.3patched).

Radiomics feature screening: The 1314 extracted radiomics features were first screened by RStudio using the least absolute shrinkage and selection operator (LASSO) logistic regression algorithm to select radiomics features with non-zero coefficients from the GTV, CTV, and GTV+CTV training cohorts, respectively. The screened features were used to build logistic regression models, and the stepwise regression method (stepAIC) was employed to test and remove the non-significant variables. A radiomics nomogram was constructed to quantify the individual probability of OLNM for the significant variables, and the relationship between the predicted and ideal values was assessed using calibration curves. The area under the curve (AUC) and receiver operating characteristic (ROC) were constructed. The decision curve index to evaluate, the area under the curve (AUC), the sensitivity, and the specificity of the radiomics characteristics were used to determine the prediction accuracy in the training and validation groups.

**3. Results**
3.1 Clinical Characteristics and Univariate Analysis

A total of 598 patients were included in this study, divided into a training group and an external validation group, of which 401 were in the training group (the mean age of patients with LNM was 60.03±9.798 years, mean age of patients without LNM was 61.35±8.009 years, the P value of independent t-test was 0.248>0.05, no statistical difference) and 197 were in the external validation group (patients with LNM The mean age of patients with LNM was 59.67±9.497 years, and the mean age of patients without LNM was 60.22±9.335 years, and the P value of independent t-test was 0.718>0.05, no statistical difference), and Table 1 shows each group’s percentage of the total population. Using a four-grid table chi-square test, the tumor stage was determined in the training group, and the P value was 0.000<0.05, suggesting a statistical difference in predicting lymph node metastasis, but the P value for the validation group was 0.071>0.05, which was not statistically significant. The remaining clinical characteristics were not significantly different between the various groups (Table 2).

<table>
<thead>
<tr>
<th>groups</th>
<th>Percentage of lymph node metastasis</th>
<th>Percentage of non-lymph node metastases</th>
<th>Total number of people</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training group</td>
<td>67 (16%)</td>
<td>334 (84%)</td>
<td>401</td>
</tr>
<tr>
<td>Validation group</td>
<td>52 (26%)</td>
<td>145 (74%)</td>
<td>197</td>
</tr>
<tr>
<td>Total number</td>
<td>119 (20%)</td>
<td>479 (80%)</td>
<td>598</td>
</tr>
</tbody>
</table>
Table 2. Clinical characteristics

<table>
<thead>
<tr>
<th>Feature</th>
<th>Training group</th>
<th>Validation group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LNM (+)</td>
<td>LNM (-)</td>
</tr>
<tr>
<td>Age, mean±SD</td>
<td>60.03±9.798</td>
<td>61.35±8.009</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>28</td>
<td>149</td>
</tr>
<tr>
<td>Female</td>
<td>39</td>
<td>185</td>
</tr>
<tr>
<td>Tumor stages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>50</td>
<td>284</td>
</tr>
<tr>
<td>T2</td>
<td>17</td>
<td>50</td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper right</td>
<td>19</td>
<td>105</td>
</tr>
<tr>
<td>Right middle</td>
<td>15</td>
<td>59</td>
</tr>
<tr>
<td>Lower right</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>Upper left</td>
<td>18</td>
<td>82</td>
</tr>
<tr>
<td>Lower left</td>
<td>10</td>
<td>67</td>
</tr>
<tr>
<td>PC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADC</td>
<td>53</td>
<td>290</td>
</tr>
<tr>
<td>SCC</td>
<td>14</td>
<td>44</td>
</tr>
</tbody>
</table>

Description: SD, standard deviation; LNM (+), lymph node metastasis; LNM (-), non-lymph node metastasis; PC, pathological classification; ADC, adenocarcinoma; SCC, squamous cell carcinomas.

3.2 Radiomics Feature Extraction and Screening

There were 401 cases in the training group and 197 cases in the external validation group, which were split in a 2:1 ratio. After outlining the region of interest (ROI) of the tumor volume, the CTV was formed by Boolean logic extrapolation, and then the extended plug-in named "Radiomics" in AccuContour software (https://www.radiomics.io/PyRadiomics.html) in AccuContour software to extract the image histology features from each ROI. The plugin automatically extracts 1314 radiomics features from the ROI of each GTV and CTV image. The GTV+CTV hybrid model was created by combining the GTV and CTV features. Finally, the eight optimal variable features were filtered out from the training set of the three models for modeling by ten-fold cross-validation using LASSO regression (Table 3).

3.3 Model Construction and Model Evaluation

By using the stepwise regression method, the features that lasso regression filtered out were added or removed one at a time. The partial regression sum of squares in the condition period introduced was then tested for significance to make sure that each independent variable in the regression model was significant (P<0.05), and
Finally, feature 92, feature 131, feature 220, feature 371, feature 902, feature 1132, feature 1205, and feature 1306 for a total of 8 features (Table 3). Three groups of multi-factor logistic regression models for GTV, CTV, and GTV+CTV were built using the various features, and the model effects were then examined using an outside validation group. Calibration curves were plotted to evaluate the agreement of predicted incidence with actual incidence according to the training group, external validation group, and mixed group models, respectively (Figure 2). Finally, three nomogram plots were used to present the built models (Figure 3). Each risk factor variable, the overall score, and the likelihood of predicted lymph node metastasis were represented on the left side of the nomogram plots, from top to bottom. The method of using the nomogram is to first find the corresponding scores according to the assigned values of each risk factor variable and then add the scores to obtain the total score, and then determine the risk of lymph node metastasis that corresponds to the total score.

### Table 3. Features and their types

<table>
<thead>
<tr>
<th>Features</th>
<th>Name</th>
<th>Feature Type</th>
<th>Image Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature92</td>
<td>Dependence variance</td>
<td>GLDM</td>
<td>Original</td>
</tr>
<tr>
<td>Feature131</td>
<td>Correlation</td>
<td>GLCM</td>
<td>Wavelet-LLH</td>
</tr>
<tr>
<td>Feature220</td>
<td>Cluster prominence</td>
<td>GLCM</td>
<td>Wavelet-LHL</td>
</tr>
<tr>
<td>Feature371</td>
<td>Dependence variance</td>
<td>GLDM</td>
<td>Wavelet-LHH</td>
</tr>
<tr>
<td>Feature902</td>
<td>Run entropy</td>
<td>GLRLM</td>
<td>Square</td>
</tr>
<tr>
<td>Feature1132</td>
<td>90th percentile</td>
<td>First order</td>
<td>Exponential</td>
</tr>
<tr>
<td>Feature1205</td>
<td>Dependence entropy</td>
<td>GLDM</td>
<td>Exponential</td>
</tr>
<tr>
<td>Feature1306</td>
<td>Large dependence high gray level emphasis</td>
<td>GLDM</td>
<td>Gradient</td>
</tr>
</tbody>
</table>

**Note:** a. Dependence Variance (DV): Measures the small difference in the magnitude of the dependency terms in the image. Correlation: Correlation is a value between 0 (no correlation) and 1 (perfect correlation), showing the linear dependence of grayscale values on their respective voxels in the GLCM. c. Cluster Prominence: It is a measure of GLCM skewness and asymmetry. Higher values indicate higher asymmetry in the mean, while lower values indicate peaks around the mean and small variations in the mean. d. Run Entropy (RE): This measures the uncertainty/randomness of the distribution of travel length and gray levels. Higher values indicate higher heterogeneity in the texture pattern. e. 90th percentile: 90th percentile. f. Entropy: Entropy specifies the uncertainty/randomness in the image values. It measures the average amount of information needed to encode the image values. g. Large Dependence High Gray Level Emphasis (LDHGLE): Measures the joint distribution of large dependencies with high gray level values. h. Gray Level Co-occurrence Matrix glcm (GLCM): Gray Level Co-occurrence Matrix, describes the texture by the spatial correlation properties of gray levels. i. Gray Level Dependence Matrix (GLDM) quantifies the gray level dependence in an image. Gray level dependence is defined as the number of connected voxels within a distance $\delta$ depending on the center voxel. j. Gray Level Travel Length Matrix (GLRLM) quantifies the gray level travel (defined as the length of the number of pixels) of consecutive pixels with the same gray level value. The $P(i,j|\theta)$, $(i,j)$ the element of the gray travel distance matrix describes the number of operations i of the gray level and the number of operations j along the length of the image (ROI) 6. k. First-Order Statistics First-order statistics describe the distribution of voxel intensities in the image region defined by the ROI using common and fundamental measures.

### 3.4 Comparison of models

The AUC of the GTV model (extracting only GTV image features) was 0.845 (95% CI: 0.801-0.888) and 0.821 (95% CI: 0.749-0.894) in the training and test sets, respectively. (95% CI: 0.749-0.894), the AUC values for the CTV model (extracting only CTV image features) were 0.834 (95% CI: 0.787-0.834) and 0.812 (95% CI: 0.780-0.912) in
the training and validation group, respectively, and the AUC values for the GTV+CTV hybrid model were 0.869 (95% CI: 0.842-0.896) and 0.906 (95% CI: 0.757-0.953) in the training and validation group, respectively (detailed results in Table 4). The higher the AUC value, the better the prediction effect of the model and the ROC curve of each model is shown in Figure 4. The AUC values of the CTV model compared with the GTV model in the external validation group did not show any advantage (Delong’s test P-value was 0.855>0.05, which was not statistically significant). The AUC values of the hybrid GTV+CTV model were higher than those of the GTV and CTV validation group models (the AUC of the hybrid model was 0.869 in the training group and 0.906 in the validation group). And the Decision curve is used to check the net benefit of each model, among which the hybrid model has the best net benefit (Figure 5). The p-values of the Delong test between the hybrid model and the remaining two groups were <0.05, which was statistically significant (the p-value of the Delong test between the hybrid and GTV groups was 0.04<0.05, and the p-value of the Delong test between the hybrid and CTV groups was 0.01<0.05).

<table>
<thead>
<tr>
<th>Radiomics Model</th>
<th>Training Group</th>
<th>Testing Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>GTV</td>
<td>0.673</td>
<td>0.912</td>
</tr>
<tr>
<td>CTV</td>
<td>0.650</td>
<td>0.940</td>
</tr>
<tr>
<td>GTV+CTV</td>
<td>0.783</td>
<td>0.838</td>
</tr>
</tbody>
</table>

Table 4. Summarizes the AUC values for the training and validation groups of the three models. The AUC of each model is >0.8, indicating that good prediction results are achieved.

4. Discussion

In this study, we constructed a model to predict occult lymph node metastasis in NSCLC based on the primary tumor focus, the peri-tumor environment of 5 mm around cancer, and the mixed features of both, and our ROC results showed that the GTV model, the CTV model and the hybrid GTV+CTV model all showed good predictive efficacy (The AUCs were 0.821, 0.812, and 0.906 in the training group, and 0.821, 0.812, and 0.906 in the validation group), respectively; the P-values of the Delong test for the hybrid model and the GTV and CTV models were less than 0.05, which were statistically significant. Additionally, the decision curves demonstrated a good net benefit, and the outcomes demonstrated that the hybrid model delivered the best prediction performance (AUC of 0.869 in the training group and AUC of 0.906 in the validation group). The Delong test performed for the GTV validation group and the CTV validation group models suggested that the AUC values were not significantly different. Moreover, the Delong tests for the hybrid model and the other two models were both statistically significant.
Two types of radiomics features, GLCM (25%) and GLDM (50%) were found to occupy the major proportion in the prediction of LNM, suggesting that researchers can focus on these two types of features in the prediction of occult LNM in lung cancer. The radiomics features in this study demonstrated better predictive efficacy compared with the clinical features. The continued predictive value of these two features for other cancers needs to be confirmed. The difference between the results of this study in the two groups of GTV and CTV was not satisfactory, considering that this result was due to the small data set of the study (only 401 cases were included in our hospital, and an additional 197 cases came from elsewhere), but the mixed training model and validation model formed by mixing GTV and CTV in the training and testing groups (which is equivalent to expanding the data set and increasing the exposure rate of the effective features) AUC predicted better than training the GTV and CTV models alone, suggesting that radiomics studies may achieve better results with larger samples. A multicenter study was conducted in this study, allowing the results to exclude single-center bias.

Many studies have been published in the literature on predicting lymph node metastasis in a variety of fields, including lung cancer, bile duct cancer, gastric cancer, and colorectal cancer, thanks to the rapid development of radiomics in recent years. Some studies have demonstrated the value of the tumor microenvironment for clinical assessment of the aggressive biological behavior of tumors, while the majority of studies concentrated on feature extraction from the primary tumor focus, ignoring the role of the environment in which the tumor cells are located in distant tumor metastasis. In contrast, radiomics can respond to macroscopic features of tumors, such as size, shape, and texture, through high-throughput, high-dimensional features, which can be used to quantitatively describe various aspects of tumors, especially for assessing regional heterogeneity. Previous studies have shown that quantitative imaging features based on histogram, texture, and shape in imaging omics can respond to information related to the tumor microenvironment. The present study precisely considered that the peritumor environment may contain some information about the tumor. First, according to the definition of CTV, there may be micro infiltrations of the tumor around the bulk of the tumor. Secondly, this study uses radiomics to describe the characteristics of the microenvironment around the tumor to determine how it relates to occult lymph node metastasis in non-small cell lung cancer and determine the best course of action for clinical patients.

Several published literatures have reported that clinical features such as tumor size, density, morphological features, and serum CEA are highly correlated with lymph node metastasis in lung cancer, and in recent years, studies of lung cancer radiomics combined with clinical features to predict lymph node metastasis have produced good predictive results. The small number of cases and inclusion of clinical features in this study, however, was thought to be related to the clinical features' poor predictive performance, so only radiomics features were included. The predictive advantage of the radiomics model was also significant, and in comparison to the radiomics model based on the bulk tumor volume, the combined model including the CTV had a higher predictive efficacy, which was thought to be related to the increased patient population.

In this study, only non-small cell lung cancer was included, and the tumor bulk volume under enhanced CT of the cases was clearly outlined. Cases with positive lymph nodes in the imaging report were excluded for the prediction of occult lymph node metastasis. Complete data from the CT report and postoperative pathological reports were gathered, and an external test group was added for a more thorough investigation of the influencing factors. Various traits were included to create three models, which, on the one hand, can be compared to show the importance and benefits of the integrated radiomics model of tumor bulk volume and CTV for predicting occult lymph node metastasis in NSCLC were better reflected.
There are still the following shortcomings in this study: Although this study is a multicenter study, the data set is less studied, the number of positive cases is low, and the sample distribution is more biased. More clinical characteristics should be added for thorough prediction to increase the predictive efficacy since we did not combine clinical characteristics for modeling prediction. Previous studies on lung cancer showed that peritumoral tissue 5-20 mm away from the tumor is closely related to tumor prognosis\textsuperscript{[29,30]}, but we only explored the effect of features around the tumor 5 mm. In this study, the radiomics features were resampled and standardized, however, the variability brought on by various scanning methods and manual image segmentation could not be eliminated, which represents a significant issue in the field of radiomics today.

5. Conclusion

In conclusion, nowadays, there is a lack of cheaper, more convenient, and safer examination methods to evaluate lymph node status in patients with early lung cancer. Therefore, it would be beneficial for patients to find occult lymph node metastases in CECT that provides more information.

References


**Figures**

**Figure 1.1**

**Figure 1.2**

**Figure 1**

Example of gross tumor volume (GTV) and clinical target volume (CTV) with 5mm external expansion on high-resolution enhanced computed tomography
Figure 2

(A) (B) (C) shows the calibration curves of GTV, CTV, and mixed models in the training group model, respectively. The MAE (Mean absolute error) is 0.025, 0.038, and 0.024 for all 500 sampling runs, respectively, and the MAE values are <0.05, indicating the model prediction calibration.
Figure 3

(D) (E) (F) shows the nomogram plots of GTV, CTV, and hybrid models, respectively, whose variables are the best features screened from radiomics.
Figure 4

(G), GTV training group; (H), GTV validation group; (I), CTV training group; (J), CTV validation group; (K), GTV+CTV mixed training group; (L), GTV+CTV mixed validation group.
Figure 5

Model-1, GTV validation group model decision curve; model-2, CTV validation group decision curve; model-3, hybrid validation group decision curve.

Supplementary Files

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

- HybridTraining.csv
- HybridValidation.csv
- TrainingCTV.csv
- TrainingGTV.csv
- ValidationCTV.csv
- ValidationGTV.csv