The CT delta-radiomics based machine learning approach in evaluating multiple primary lung carcinoma

Yanqing Ma  
Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College

Jie Li  
Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College

Xiren Xu  
Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College

Yang Zhang  
Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College

Yi Lin  (✉️ Linyi197001@163.com )  
Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College

Research Article

Keywords: Lung, carcinoma, multiple, computed tomography, radiomics

Posted Date: June 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1728327/v1

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

Object:
To evaluate the difference between multiple primary lung carcinoma (MPLC) and solitary primary lung carcinoma (SPLC) by delta-radiomics based machine learning algorithms in CT images.

Methods

A cohort of 1094 patients containing 268 MPLCs and 826 SPLCs were enrolled for radiomic study. After the segmentation of volume of interest, the radiomic features were automatically calculated. The patients were divided into the training set and validation set with a random proportion of 7:3. After the methods of feature selection, the relevant classifiers were constructed by the machine learning algorithms of Bayes, forest, k-nearest neighbor, logistic regression, support vector machine, and decision tree. The relative standard deviation (RSD) was calculated and the classification model with minimal RSD was chosen for delta-radiomics analysis to explore the variation of tumor during follow-up surveillance in the cohort of 225 MPLCs and 320 SPLCs. According to the different follow-up duration, three groups were divided into group A (3–12 months), group B (13–24 months), and group C (25–48 months). Then the corresponding delta-radiomics classifiers were developed to evaluate MPLCs. The area under the receiver operator characteristic curve (AUC) with 95% confidence interval (CI) was quantified to evaluate the efficiency of the model.

Results

To radiomics analysis, the forest classifier (FC-radio) with the minimal RSD showed the better stability was chosen with AUCs of 0.840 (95%CI, 0.810–0.867) and 0.670 (95%CI, 0.611–0.724) in the training and validation set. The AUCs of the forest classifier based on delta-radiomics (FC-delta) were higher than those of FC-radio. In addition, with the extension of follow-up duration, the performance of FC-delta in Group C were the best with AUCs of 0.998 (95%CI, 0.993-1.000) in the training set and 0.853 (95%CI, 0.752–0.940) in the validation set.

Conclusions

The machine-learning approach based on radiomics and delta-radiomics helped to differentiate SPLCs from MPLCs. The FC-delta with longer follow-up duration better differentiated SPLCs from MPLCs.

Introduction

Lung carcinoma is the most common cause of cancer-related death in china and is also a major health challenge worldwide[1]. The frequency and detection of lung carcinoma has descended gradually in the
US, while it is remarkably increased in China over recent years\cite{2}. Patients who survive one occurrence of non-small-cell lung carcinoma are at high risk of a second malignancy\cite{3}. Cases of multiple primary lung carcinoma (MPLCs) are increasing, mainly leading to the improved diagnostic strategies, surveillance modalities, and the aging population\cite{4}. The diagnostic criteria of MPLCs was firstly established by Martini and Melamed in the year of 1975\cite{5} and was renewed by the American College and Chest Physicians (ACCP) in 2007\cite{6}. The MPLC is classified into synchronous phenotype (sMPLC) when the second lung carcinoma was simultaneously diagnosed within 2 years after the primary lesion, and metachronous phenotype (mMPLC) when it was separately diagnosed more than 2 years after the initial surgery\cite{7}. Duchateau et al. firstly indicated that 25% of patients accompanied with MPLCs, and patients with and without MPLCs had different growth habits\cite{8}.

Evaluation the MPLCs from SPLCs by empiric radiological experience is difficult\cite{7}. Further analysis is therefore needed to comprehend more clearly. Radiomics converts the traditional radiological images into a large amount of minable high-dimensional data to explore the potential imaging biomarkers and support decision making\cite{9}. And the delta-radiomics is the change of radiomic features after treatment or surveillance\cite{10}. Undoubtedly, a delta-radiomics approach will have a growing impact on distinction the MPLCs and SPLCs, which will enable optimized management of patients with MPLCs. To best of our knowledge, there is no study focused on the delta-radiomics difference between MPLCs and SPLCs. The purpose of our study is to evaluate the delta-radiomic influence of MPLCs on prognosis to help us better understand their difference with SPLCs.

**Materials And Methods**

This retrospective study was approved by the Institutional Review Board of our hospital (NO. 2020QT108), which waived the informed consent of all patients.

**Patients screening**

This retrospective study enrolled 1094 patients, which were pathologically diagnosed as lung adenocarcinoma after 6 years follow-up surveillance, including 826 patients with SPLCs and 268 MPLCs, between January 2014 to December 2020. Among these patients there were 320 SPLCs patents and 225 MPLCs patents with regular surveillance were incorporated for delta-radiomics analysis (Fig. 1). The inclusion criteria were as follows: (1) tumors were classified to be MPLCs according to the criteria of the 2nd edition of ACCP evidence-based clinical practice guidelines\cite{6} (Table 1); (2) patients had only one primary lung tumor at the time cut-off of inclusion were classified to be SPLCs; (3) patients were pathologically proved to be minimally invasive (MIA) or invasive adenocarcinoma (IAC) of lung; (4) patients underwent CT examinations with the same protocol. The exclusion criteria were as follows: (1) patients were pathologically confirmed to be atypical adenomatous hyperplasia, in situ adenocarcinoma, or pulmonary squamous carcinoma; (2) patients was pathologically confirmed by needle biopsy; (3) patients were treated with the methods of radiation, chemotherapy, or radio-chemotherapy.
Table 1
The criteria to diagnose MPLCs according to ACCP

<table>
<thead>
<tr>
<th>MPLCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same histology, anatomically separated</td>
</tr>
<tr>
<td>Carcinomas in different lobes</td>
</tr>
<tr>
<td>And no N2,3 involvement</td>
</tr>
<tr>
<td>And no systemic metastases</td>
</tr>
<tr>
<td>Same histology, temporally separated</td>
</tr>
<tr>
<td>≥ 4-yr interval between carcinomas</td>
</tr>
<tr>
<td>And no systemic metastases from either carcinoma</td>
</tr>
<tr>
<td>Different histology type</td>
</tr>
<tr>
<td>Different histology type</td>
</tr>
<tr>
<td>Or different molecular genetic characteristics</td>
</tr>
<tr>
<td>Or arising separately from foci of carcinoma in situ</td>
</tr>
</tbody>
</table>

CT examination and Volume of interest segmentation

All the patients underwent CT unenhanced examinations in Somanton Definition AS 64 or 128 CT (Siemens Medical Solutions, Germany). The scan parameters were as follows: tube voltage, 120 kVp; tube current, 200mA; rotation speed, 0.75s; beam pitch, 1.375; pixel matrix, 512*512; detector collimation, 64*0.625mm; slice thickness, 2.0mm; reconstruction interval, 2.0mm; width of lung window, 1500HU; level of lung window, -600HU.

The volume of interest (VOI) of tumor was depicted in software of “ITK-snap 3.8.0” (http://www.itksnap.org/pmwiki/) by two radiologists with 10 and 12 years of experience, manually (Fig. 2a, b). Then, the radiomic features were automatically calculated in software of “A.K. 3.0.0” (GE Healthcare) after steps of preprocessing involved resampling images to be 1.0mm at X/Y/Z space, reducing the image noise by a method of Gaussian, and discretizing the gray level to the range of 1 to 32. The intra-class correlation coefficients (ICCs) of radiomic features from two radiologists were calculated to evaluate the agreement between different observers. The radiomic features with ICCs greater than 0.75 were selected and the mean values of two radiologists were calculated for further analysis.

Radiomics and delta-radiomics analysis

Prior to radiomic analysis, the steps of excluding variables with zero variance, replacing abnormal values by median values, and standardization were adopted to normalize radiomic features. Then the cohort was randomly divided into the training set and validation set with a random proportion of 7:3. In order to eliminate the influence of unbalanced sample size, the way of synthetic minority over-sampling technique
(SMOTE) was carried out\cite{11}. After the methods of analysis of variance, correlation analysis with a threshold of 0.7, and gradient boosting decision tree (GBDT), the optimal radiomic features were extracted. Ultimately, the corresponding machine learning based classifiers including Bayes, forest, k-nearest neighbor, logistic regression, support vector machine, and decision tree algorithms were developed to identify MPLCs and SPLCs. The relative standard deviation (RSD) was calculated and the classification model with minimal RSD was chosen for further analysis. The area under the curve (AUC) with 95% confidence interval (95%CI) of receiver operator characteristic curve (ROC) was quantified to evaluated the efficiency of the machine learning based classifiers.

With consideration of the different progression during regular follow-up surveillance between MPLCs and SPLCs, the delta-radiomics was utilized. The delta-radiomics was defined as the change of radiomic features between baseline and follow-up surveillance, including Group A (3–12 months), Group B (13–24 months), and Group C (25–48 months). The equation of delta-radiomics was: \( \frac{(\text{follow-up radiomics} - \text{baseline radiomics})}{\text{follow-up interval}} \). The specific information of radiomics and delta-radiomics analysis were listed in Supplementary Material.

**Statistics**

The general clinical characteristics were analysis by software of “SPSS 22.0“ with methods of student’s t-test or chi-square test. The methods of radiomic feature selection including variance, correlation analysis, GBDT, and machine learning algorithms were performed by the software of “Python 3.5”. The ROC curve was delineated by the software of “MedCalc 15.8”. A \( p \)-value less than 0.05 indicates statistical significance.

**Results**

**Patient’s general information**

There were 1094 patients with 268 MPLCs and 826 SPLCs. The general information of all patients were listed in Table 2. The general information included gender, age, location, and pathology. The variables of gender \( (p = 0.279) \) and age \( (p = 0.575) \) had no statistical significance, while the variables of location \( (p < 0.05) \) and pathology \( (p < 0.05) \) showed the significant difference. The tumors of SPLCs were more likely to locate in the right lung than these of MPLCs (62.2% vs. 60.1%). The pathological type of MPLCs was easier to be MIA (45.9% vs. 32.7%), while that of SPLCs was more prone to be IAC (67.3% vs. 54.1%).
Table 2
Patients’ general information

<table>
<thead>
<tr>
<th></th>
<th>MPLCs (n = 268)</th>
<th>SPLCs (n = 826)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>0.279</td>
</tr>
<tr>
<td>Female (%)</td>
<td>168 (62.7%)</td>
<td>487 (59.0%)</td>
<td></td>
</tr>
<tr>
<td>Male (%)</td>
<td>100 (37.3%)</td>
<td>339 (41.0%)</td>
<td></td>
</tr>
<tr>
<td>Age (mean ± standard)</td>
<td>57.4 ± 10.6</td>
<td>56.9 ± 12.9</td>
<td>0.575</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Right lung</td>
<td>161 (60.1%)</td>
<td>514 (62.2%)</td>
<td></td>
</tr>
<tr>
<td>Superior lobe</td>
<td>76 (28.4%)</td>
<td>298 (36.1%)</td>
<td></td>
</tr>
<tr>
<td>Middle lobe</td>
<td>40 (14.9%)</td>
<td>54 (6.5%)</td>
<td></td>
</tr>
<tr>
<td>Inferior lobe</td>
<td>45 (16.8%)</td>
<td>162 (19.6%)</td>
<td></td>
</tr>
<tr>
<td>Left lung</td>
<td>107 (39.9%)</td>
<td>312 (37.8%)</td>
<td></td>
</tr>
<tr>
<td>Superior lobe</td>
<td>71 (26.5%)</td>
<td>199 (24.1%)</td>
<td></td>
</tr>
<tr>
<td>Inferior lobe</td>
<td>36 (13.4%)</td>
<td>113 (13.7%)</td>
<td></td>
</tr>
<tr>
<td>Pathology</td>
<td></td>
<td></td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>MIA</td>
<td>123 (45.9%)</td>
<td>270 (32.7%)</td>
<td></td>
</tr>
<tr>
<td>IAC</td>
<td>145 (54.1%)</td>
<td>556 (67.3%)</td>
<td></td>
</tr>
</tbody>
</table>

Radiomics analysis

The radiomic features between the patients of MPLCs and SPLCs were differentiated. There were 27 radiomic features remained after feature selection of GBDT (Fig. 3) and the machine learning based classifiers including Bayes, forest, k-nearest neighbor, logistic regression, support vector machine, and decision tree were constructed in the training set and confirmed in the validation set. The forest classifier of radiomics (FC-radio) with minimal RSD of 1.82 was chosen for further analysis (Supplementary Material, Table 1). The specific AUC values of six machine-learning algorithms from 100 Bootstrap replication in the training set were listed in Supplementary Material, Table 2. The AUC of this FC-radio in the training set was 0.840 (95%CI, 0.810–0.867) and that of the validation set was 0.670 (95%CI, 0.611–0.724). The low discrimination efficiency of this model indicates that the radiomic difference between tumors of SPLCs and MPLCs was inconspicuous.

Delta-radiomics analysis

Regardless of the poor efficiency of the FC-radio in distinguishing MPLCs and SPLCs, a further forest machine learning algorithm of delta-radiomics (FC-delta) was conducted. Depending on the different
duration of follow-up, we divided the patients into four groups: group A with a follow-up intervals of 3–12 months (105 MPLCs vs. 145 SPLCs), group B with a follow-up intervals of 13–24 months (68 MPLCs vs. 96 SPLCs ), and group C with a follow-up intervals of 25–48 months (52 MPLCs vs. 79 SPLCs).

The AUC of FC-delta in group A was 0.972 (95%CI, 0.951–0.989) in the training set and was 0.798 (95%CI, 0.704–0.892) in the validation set. The AUC of FC-delta in group B was 0.989 (95%CI, 0.978–0.997) in the training set and was 0.821 (95%CI, 0.708–0.915) in the validation set. The AUC of FC-delta in group C was 0.998 (95%CI, 0.993-1.000) in the training set and was 0.853 (95%CI, 0.752–0.940) in the validation set. With the extension of follow-up intervals, the difference between MPLCs and SPLCs was more obvious (Fig. 4).

Discussions

We reported a single-institution experience on radiomic differentiation on MPLCs from SPLCs, especially emphasis on their long-term variation. The reported incidence of synchronous lung carcinoma is variably between 0.2–20%\[12\]. In our study, the incidence of MPLCs was 24.5%, which was slightly higher than the reported incidence. It may be related to the universality of chest computed tomography screening programs. There were no statistical difference between the characteristics of gender and age. Slightly different from the outcome in the past study, female gender and smoke free statue were more frequent in MPLCs\[^7\]. And the MPLCs were more easier to the pathological type of MIA compared with SPLCs (45.9% vs. 32.7%, \(p<0.05\)). This result supported the previous view that the proportion of MIA and adenocarcinoma in situ is high in sMPLCs\[^13\]. This also may be due to timely detection of MPLCs at a early stage with regular surveillance for the first primary carcinoma\[^14\]. However, present clinical and traditional methods are unable to understand the different evolution between MPLCs and SPLCs. Thus, studies of novel factors that differ significantly between patients with MPLCs and SPLCs are necessary vehicles for identifying subtleties in two diseases.

Patients with first primary cancer remain at risk of developing a secondary tumor at a distant site through metastasis via the lymphatic or circulatory system\[^15\]. Second primary carcinomas showed specific associations with the first one and their nature course were not the same\[^16\]. The etiology of MPLCs is ambiguous\[^17\]. Our radiomics analysis between MPLCs and SPLCs found of interest that radiomics could identify the difference between two groups with the AUC of 0.840 (95%CI, 0.810–0.867) in the training set and that of 0.670 (95%CI, 0.611–0.724) in the validation set. Nevertheless, this low discrimination efficiency was insufficient to supplied accurate information to better understand the difference between MPLCs and SPLCs. To best of our knowledge, it is the first article focused on the distinction between MPLCs and SPLCs from the point view of radiomics and delta-radiomics.

The crucial challenge regarding MPLCs is what they differed with SPLCs in the course of development, on which both the treatment strategies and prognosis are based\[^18\]. The possible of difference between MPLCs and SPLCs should always be considered during the follow-up surveillance, which determines the
subsequent management strategy\textsuperscript{[19]}. It has conclusively been suggested that the overall survival of MPLCs was better than SPLCs with intrapulmonary metastasis\textsuperscript{[20]}. Therefore, a delta-radiomics approach studied the variation of radiomic features during baseline examination and follow-up duration\textsuperscript{[21]}. The AUC of FC-delta of group C was the highest both in the training set (0.998 vs. 0.989 and 0.972) and the validation set (0.853 vs. 0.821 and 0.798). With the extension of follow-up intervals, the difference between MPLCs and SPLCs was more obvious. The literature on survival difference between sMPLCs and SPLCs has quantified and highlighted that the prognosis of sMPLCs was poorer and resembled that of SPLCs of a higher stage\textsuperscript{[22]}. Our results suggested that the nature course of two diseases was inconsistent and the delta-radiomics could better distinct the MPLCs and SPLCs than radiomics. We firstly reported the difference of two diseases in terms of both radiomics and delta-radiomics to help us make decision on individual therapy and predict the prognosis of diseases.

Our present study has several limitations. Currently, there are no definitive guideline for the diagnosis and treatment of MPLCs. In 2003, the American College of Chest Physicians (ACCP) developed a new diagnostic criteria for MPLCs with evaluations of lymphatic and systemic metastasis and the interval between mMPLC was extended to at least 4 years\textsuperscript{[23]}. Antakli et al. revised the criteria of Martini and Melamed by adding DNA ploidy validation for distinction\textsuperscript{[24]}. However, they have not widely applied to clinical practice due to its disadvantages of expensive, time consuming, and low sensitivity. Hence, we adopted the most cited criteria of the 2nd edition of ACCP in our research. Second, the MPLCs can be subdivided into mMPLCs and sMPLCs. Due to the limitation of incidence and sample size, we performed a general analysis of MPLCs which may lead to a biased result. Third, we only enrolled the cohort with pathological types of MIA and IAC to analysis and neglected other pathological types of lung carcinomas. The MPLCs and SPLCs with pathological types of adenocarcinoma in situ, squamous carcinoma\textsuperscript{[25]}, and so on should further be studied after collecting enough cases.

**Conclusion**

In conclusion, our study revealed that the approaches of radiomics and delta-radiomics help to differentiate MPLCs and SPLCs. The radiomic difference between SPLCs and MPLCs was faint and the delta-radiomics better differentiate these patients. Moreover, with the extension of follow-up duration, the delta-radiomics difference between SPLCs and MPLCs appeared more distinctly.

**Declarations**

**Ethics approval and consent to participate:** This retrospective study was approved by the Medical Ethics Committee of Zhejiang Provincial People’s Hospital (NO. 2020QT108) and in conformity to the Declaration of Helsinki. The informed consent was waived for this retrospective study by the Medical Ethics Committee of Zhejiang Provincial People’s Hospital (NO. 2020QT108).

**Consent for publication:** NA.
Availability of data and materials: The datasets used and analyzed in this article is available from the corresponding author on reasonable request. The code used in this study is available at GitHub (https://github.com/mayq1988/GGN).

Competing interests: No competing interests.

Funding: The study was supported by the fund of Medical and Health Research Project of Health Commission of Zhejiang Province (No. 2022492695).


Acknowledgements: not applicable.

References


Figures
2075 patients underwent pulmonary surgeries

1094 patients of lung adenocarcinoma

Radiomics analysis

268 MPLCs
826 SPLCs

Delta-Radiomics analysis

225 MPLCs
320 SPLCs

Group A (3-12 months)
105 MPLCs vs. 145 SPLCs

Group B (13-24 months)
68 MPLCs vs. 96 SPLCs

Group C (25-48 months)
52 MPLCs vs. 79 SPLCs

Figure 1

The flow diagram of patients selection.
Figure 2

The VOI of tumor was manually depicted in the software of “ITK-snap”.

Figure 3

The heatmap of radiomics analysis after feature selection of GBDT, and there were 27 radiomic features selected.
Figure 4

The comparison of AUCs of FC-radio and FC-delta of Group A, B, and C in the training and validation set.

Supplementary Files

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

- SupplementMaterial.doc