A Nomogram Based on CT Deep Learning Signature: A Potential Tool for the Prediction of Overall Survival in Resected Non-Small Cell Lung Cancer Patients

Purpose: To develop and further validate a deep learning signature-based nomogram from computed tomography (CT) images for prediction of the overall survival (OS) in resected non-small cell lung cancer (NSCLC) patients.

Patients and Methods: A total of 1792 deep learning features were extracted from non-enhanced and venous-phase CT images for each NSCLC patient in training cohort (n=231). Then, a deep learning signature was built with the least absolute shrinkage and selection operator (LASSO) Cox regression model for OS estimation. At last, a nomogram was constructed with the signature and other independent clinical risk factors. The performance of nomogram was assessed by discrimination, calibration and clinical usefulness. In addition, in order to quantify the improvement in performance added by deep learning signature, the net reclassification improvement (NRI) was calculated. The results were validated in external validation cohort (n=77).

Results: A deep learning signature with 9 selected features was significantly associated with OS in both training cohort (hazard ratio [HR]=5.455, 95% CI: 3.393–8.769, P<0.001) and external validation cohort (HR=3.029, 95% CI: 1.673–5.485, P=0.004). The nomogram combining deep learning signature with clinical risk factors of TNM stage, lymphatic vessel invasion and differentiation grade showed favorable discriminative ability with C-index of 0.800 as well as a good calibration, which was validated in external validation cohort (C-index=0.723). Additional value of deep learning signature to the nomogram was statistically significant (NRI=0.093, P=0.027 for training cohort; NRI=0.106, P=0.040 for validation cohort). Decision curve analysis confirmed the clinical usefulness of this nomogram in predicting OS.

Conclusion: The deep learning signature-based nomogram is a robust tool for prognostic prediction in resected NSCLC patients.

Keywords: deep learning, non-small cell lung cancer, prognosis, nomogram

Introduction

Non-small cell lung cancer (NSCLC), the most frequent cause of cancer-related mortality, accounts for approximately 85% of lung cancer, and surgical resection is the preferred treatment for stage I–IIIA NSCLC. However, overall survival (OS) still varies even though the tumor is completely resected. Currently, the tumor-node-metastasis (TNM) staging system is the routine method to estimate...
prognosis. However, the survival outcomes could be dramatically different among patients with the same TNM tumor stage. Therefore, there is an urgent need to develop a new prognosis estimation method for achieving personalized precision medicine.

Computed tomography (CT), a routinely used imaging technique for disease diagnosis, provides great opportunities for personalized medicine. Radiomics, a recently emerging method for characterizing tumor heterogeneity with high-dimensional quantitative features extracted from medical images, has shown promising results in evaluation of tumor treatment response and prognosis. Most previous studies have relied on hand-crafted features which encode our prior knowledge on the data. However, hand-crafted features are low-order and susceptible to noise, and may not be adequate for unveiling the characteristics of tumors.

Currently, deep learning based on a neural network structure has shown great potential in medical images. It can automatically extract high-level features from pixel images for tumor classification, segmentation and detection. Hosny et al reported that deep learning could be used to improve risk stratification for NSCLC patients. However, a deep learning signature they developed for binary prediction of 2-year OS relied on arbitrary thresholds and cannot predict OS as a continuous variable. Moreover, they failed to further develop a model which might be useful for clinical practice.

Therefore, we aimed to construct and validate a signature-based nomogram to predict OS as a continuous variable and to evaluate the incremental prognostic value of the deep learning signature from preoperative CT images for individual OS prediction in patients with resected NSCLC.

Patients and Methods

Patients

Ethical approval for this study was obtained from the Ethics Committees of Guangdong Provincial People’s Hospital (Hospital 1) and Zhujiang Hospital (Hospital 2), and the review board exempted the acquisition of informed consent because this was a retrospective study. Confidentiality of patient data was maintained anonymously, and all procedures performed in our study were in compliance with the Declaration of Helsinki. Inclusion criteria were described below: (a) patients pathologically diagnosed NSCLC; (b) both preoperative non-enhanced and venous-phase CT scanning were available. Exclusion criteria were described below: (a) missing preoperative non-enhanced or venous-phase CT images; (b) multiple primary carcinomas or concurrent malignancies; (c) patients received chemotherapy or radiotherapy before surgery; (d) clinical characteristics were incomplete; (e) Patients with obvious artifacts in CT images. Finally, a total of 308 patients were included in the study. Of 308 patients, 231 patients enrolled between January 2007 and August 2014 from Hospital 1 served as the training cohort, and 77 patients enrolled between March 2010 and December 2015 from Hospital 2 served as the validation cohort.

Clinical Characteristics and Follow-Up

Baseline clinical information was retrieved from the institutional database for medical records, including age, gender, smoking status, TNM stage, the status of lymphatic vessel invasion, differentiation grade, pathological type and location of tumor.

The primary outcome was OS, defined as the time from the date of preoperative CT scan to the date of death from any cause (event) or last follow-up (censored). Patients were followed up at least 3 years postoperatively. The investigators were blinded to the clinical variables and patients’ outcome.

Image Acquisition

All patients underwent non-enhanced and venous-phase chest CT scans. The parameters are listed in Appendix A1.

Tumor Segmentation

All preoperative non-enhanced and venous-phase chest CT images were obtained from the picture archiving and communication system (PACS). A radiologist with 8-year experience in interpreting chest CT images manually delineated the region of interest (ROI) along the maximum cross-sectional border of the tumor using ImageJ (National Institutes of Health, Bethesda, MD). This procedure was performed on none-enhanced and venous-phase chest CT images, respectively.

To analyze the interobserver reproducibility of features, another radiologist with 5-year experience in chest CT interpretation randomly chose 50 cases for segmentation. Then, interclass correlation coefficients (ICCs) were calculated, and features with ICCs larger than 0.75 were considered highly reproducible.

Image Pre-Processing

In order to correct the influence of different CT devices on feature extraction, the following pre-processing steps were performed before feature extraction. First of all, all CT images were normalized to a consistent voxel size of 0.5×0.5×1 mm³.
Then, the gray values of the segmented ROI were converted into the range (0, 300) using linear transformation.

**Deep Learning Feature Extraction**

A deep architecture, ResNet-18, was applied to extract deep learning features from both preoperative non-enhanced and venous-phase chest CT images. ResNet-18 was pretrained on the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC-2012) dataset, and all weights of the network were predetermined, including 17 convolution layers (Conv1-17) and 1 fully connected layer (fc18).

CT images were grayscale images (single-channel images), while three-channel and 224×224 pixel² RGB input images (224×224×3) were required for the pretrained ResNet-18 on the ILSVRC-2012 dataset. Therefore, the following two steps were taken to meet the requirements of the pretrained ResNet-18 inputs images. First, the segmented tumor region was read, and then a bounding box covering the whole tumor area was cropped (size, 224×224 pixels). Deep learning features would be extracted from the output of Conv3_2, conv4_2, conv5_2 of pretrained ResNet-18. The toolbox of deep learning feature extraction was implemented in MATLAB 2019a (http://www.mathworks.com/help/deeplearning/ug/extract-image-features-using-pretrained-network.html). The flow chart of feature extraction process is shown in Figure 1.

**Deep Learning Feature Selection and Signature Development**

The following steps were performed for feature selection. Firstly, robust features with ICCs >0.75 were chosen and normalized by Z-score transformation. Secondly, the least absolute shrinkage and selection operator (LASSO) Cox regression was applied to select prognosis-related features from the robust features. Finally, the deep learning signature was built by fitting the selected features with the Cox proportional hazards model. A risk score was calculated as a linear combination of all the selected features with their weighting coefficients.

**Evaluation and Validation of the Signature**

Signature performance was evaluated in the training cohort and then validated in the external validation cohort. The association of the deep learning signature with OS was assessed by univariate Cox regression analysis. Patients were stratified into high- and low-risk group using the median of the risk score as the cut-off point. Survival curves were plotted using the Kaplan–Meier method, and the Log rank test was used to compare the difference of survival curves between high- and low-risk groups. The concordance index (C-index) was calculated to assess the predictive discrimination of the signature. Predictive accuracy was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC) for 3-year OS.

We also assessed whether the deep learning signature could further stratify patients within subgroups as defined by: age (≤60 or >60 years), gender (female or male), smoking status (yes or no), TNM stage (I–III), lymphatic vessel invasion (yes or no), and differentiation grade (well, moderate, poor).

**Construction and Assessment of the Nomogram**

Univariate and multivariate Cox regression analyses were performed to identify the prognostic factors. To assess the incremental prognostic value of the deep learning signature for individual OS prediction, a deep learning signature-based model and a clinical model were built. Variables included in the deep learning signature-based model were selected via the stepwise selection method by minimizing Akaike’s information criterion (AIC) and are expressed as a nomogram. The clinical model was built with only clinical risk factors.

We assessed the performances of the nomogram and the clinical model in both the training and the external validation cohorts. Predictive discrimination was evaluated using C-index, and predictive accuracy was measured using AUC for 3-year OS estimation. A calibration curve was presented to evaluate the concordance between the observed outcome and the estimated survival probability. To quantify the improvement of deep learning signature for the nomogram, the net reclassification improvement (NRI) was calculated. Finally, we employed decision curve analysis to estimate the clinical usefulness (net benefit over a range of risk thresholds).

**Statistical Analysis**

The Chi-square test or the Mann–Whitney U-test as appropriate was used to compare the differences of baseline clinical information between the training and the external validation cohorts.

Statistical analyses were performed with R programming language (R version 3.3.1; http://www.R-project.org). The R packages employed included “glmnet”, “survival”, and “timeROC”. All statistical tests were two-sided with a significance level of 0.05.
**Results**

**Clinical Characteristics and Follow-Up**

The clinical characteristics for the training and external validation cohorts are listed in Table 1. No statistically significant difference was detected between two cohorts with regard to gender, smoking status, lymphatic vessel invasion or tumor location (P=0.109–0.785) except for the factors of age, TNM stage, differentiation grade, follow-up time and pathological type (P≤0.028).

The maximum follow-up times were 116 months in the training cohort and 119 months in the external validation cohort.

**Deep Learning Feature Selection and Signature Construction**

In our study, 896 deep learning features were extracted from non-enhanced and venous-phase CT images, respectively. Therefore, there were a total of 1792 deep learning features for each patient, and 435 features with ICCs>0.75 showed good reproducibility. Nine features were finally selected from the robust features by LASSO Cox regression (Figure 2), and the signature was constructed using the selected features weighted by each coefficient (see Appendix A2).

**Evaluation and Validation of the Signature**

Deep learning signature was significantly associated with OS in the training cohort (hazard ratio [HR]=5.455, 95% CI: 3.393–8.769, P<0.001) and the external validation cohort (HR=3.029, 95% CI: 1.673–5.485, P=0.004) (Figure 3). Patients were classified into high- and low-risk groups with the cut-off point (median risk score=−0.116). Patients with lower risk scores had better OS (median (interquartile range [IQR]), 1606 [1044–2082] days for the
Table 1 Clinical Characteristics of Patients with Resectable NSCLC in Training and Validation Cohort

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Training Cohort (n=231)</th>
<th>Validation Cohort (n=77)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>63 (26–85)</td>
<td>60 (28–77)</td>
<td>0.028*</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>144 (62.3%)</td>
<td>50 (64.9%)</td>
<td>0.785**</td>
</tr>
<tr>
<td>Female</td>
<td>87 (37.7%)</td>
<td>27 (35.1%)</td>
<td></td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>158 (68.4%)</td>
<td>51 (66.2%)</td>
<td>0.833**</td>
</tr>
<tr>
<td>Yes</td>
<td>73 (31.6%)</td>
<td>26 (33.8%)</td>
<td></td>
</tr>
<tr>
<td>TNM stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>158 (68.4%)</td>
<td>37 (48.1%)</td>
<td>0.004**</td>
</tr>
<tr>
<td>II</td>
<td>21 (9.1%)</td>
<td>14 (18.2%)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>52 (22.5%)</td>
<td>26 (33.7%)</td>
<td></td>
</tr>
<tr>
<td>Lymphatic vessel invasion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>199 (86.1%)</td>
<td>70 (90.9%)</td>
<td>0.373**</td>
</tr>
<tr>
<td>Yes</td>
<td>32 (13.9%)</td>
<td>7 (9.1%)</td>
<td></td>
</tr>
<tr>
<td>Differentiation grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well</td>
<td>14 (6.1%)</td>
<td>11 (14.3%)</td>
<td>0.001**</td>
</tr>
<tr>
<td>Moderate</td>
<td>168 (72.7%)</td>
<td>36 (46.8%)</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>49 (21.2%)</td>
<td>30 (38.9%)</td>
<td></td>
</tr>
<tr>
<td>Follow-up time (days)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>1314 (781.5–2029.5)</td>
<td>1940 (1296–2268)</td>
<td>0.007*</td>
</tr>
<tr>
<td>Maximum</td>
<td>3520</td>
<td>3611</td>
<td></td>
</tr>
<tr>
<td>Pathological type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>47 (20.4%)</td>
<td>6 (7.8%)</td>
<td>4.97e-10**</td>
</tr>
<tr>
<td>ADC</td>
<td>183 (79.2%)</td>
<td>57 (74.0%)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.4%)</td>
<td>14 (18.2%)</td>
<td></td>
</tr>
<tr>
<td>Location of tumor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>17 (7.4%)</td>
<td>11 (14.3%)</td>
<td>0.1091**</td>
</tr>
<tr>
<td>Peripheral</td>
<td>214 (92.6%)</td>
<td>66 (85.7%)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Unless otherwise specified, data are expressed as median for continuous variables, or number (%) for categorical variables. *p value was calculated with the Mann–Whitney test, **p value was calculated with the Pearson χ² test.

Abbreviations: NSCLC, non-small cell lung cancer; TNM, tumor, node, metastasis; SC, squamous carcinoma; ADC, adenocarcinoma.

Patients in each TNM stage were successfully divided into high- and low-risk groups with the median risk score based on the training cohort (P <0.05). The result of stratification analysis based on other clinical risk factors are shown in Appendix A3, A4 and Figure A2.

Construction and Assessment of Nomogram

Univariate and multivariate Cox regression analyses are shown in Appendix A5. A deep learning signature-based model, comprising the deep learning signature and clinical risk factors of TNM stage, lymphatic vessel invasion and differentiation grade identified in univariate analysis, was selected based on a minimized AIC=466.71 (Appendix A6) and was expressed as a nomogram (Figure 5). A clinical model was constructed with only the clinical risk factors: TNM stage, lymphatic vessel invasion and differentiation grade.
The nomogram showed better discrimination performance in predicting OS than the clinical model in both the training cohort (C-index [95% CI], 0.800 [0.746, 0.855] for the nomogram; 0.786 [0.726, 0.847] for the clinical model, \( P<0.001 \)) and the external validation cohort (C-index [95% CI], 0.723 [0.634, 0.813] for the nomogram, 0.679 [0.578, 0.778] for the clinical model, \( P<0.001 \)). The ROCs for 3-year OS for the nomogram and clinical model in both

![Diagram](image)

**Figure 3** Kaplan–Meier survival analyses according to the deep learning signature for patients in training and external validation cohorts. A significant association of the signature with the OS was shown in the training cohort, which was then confirmed in the external validation cohort. **Abbreviation:** OS, overall survival.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Training Cohort</th>
<th>Validation Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High-Risk Group</td>
<td>Low-Risk Group</td>
</tr>
<tr>
<td>No. of patients</td>
<td>115</td>
<td>116</td>
</tr>
<tr>
<td>3-year OS (days)</td>
<td>[1177 (620–1897)]</td>
<td>[1606 (1044–2082)]</td>
</tr>
<tr>
<td>Median(^1)</td>
<td>1317</td>
<td>1598</td>
</tr>
<tr>
<td>Mean</td>
<td>1177</td>
<td>1606</td>
</tr>
<tr>
<td>No. of Death</td>
<td>[9 (7.83)]</td>
<td>[3 (2.59)]</td>
</tr>
<tr>
<td>At 1-year</td>
<td>[20 (17.39)]</td>
<td>[4 (3.45)]</td>
</tr>
<tr>
<td>At 2-year</td>
<td>[27 (23.48)]</td>
<td>[6 (5.17)]</td>
</tr>
</tbody>
</table>

**Table 2** OS and Death Rate in High-Risk and Low-Risk Groups

**Note:** \(^1\)Data are interquartile ranges. **Abbreviation:** OS, overall survival.
cohort are presented in Figure 6A and B. The calibration curve showed satisfactory concordance between the observed outcome and estimated survival probability in both cohorts (Figure 6C and D). The additional value of the deep-learning signature to the nomogram was statistically significant (NRI [95% CI], 0.093 [0.004, 0.192], P=0.027 for the training cohort; NRI [95% CI], 0.106 [0.001, 0.236], P=0.040 for the validation cohort). And a decision curve analysis showed that the nomogram had a good overall net benefit (Figure 6E).

**Discussion**

In the present study, we developed a deep learning signature, which successfully stratified NSCLC patients into high- and low-risk groups. Moreover, the deep learning signature-based nomogram, comprising the clinical risk factors of the TNM stage, lymphatic vessel invasion and differentiation grade, has a better performance than the clinical model in the survival estimation. This indicates the incremental prognostic value of the deep learning signature to the TNM stage and other clinical risk factors for individual OS prediction. A deep learning signature-based nomogram could be used as a valuable tool for clinical decision-making.

TNM stage is widely used in clinical practice for cancer treatment decision-making, while it may limit the clinical usefulness for personalized medicine. Meanwhile, a single risk factor without any modelling is unable to offer a comprehensive assessment of postoperative outcome. Thus, a statistical model with multiple risk factors is necessary. Shi et al constructed a nomogram based on clinical risk factors to predict OS for NSCLC patients, showing a better discrimination ability as well as a better net benefit than the TNM stage alone. In our study, it is worth noting that a TNM stage-based clinical model showed discrimination with lower C-index than did the deep learning signature-based nomogram. This result is consistent with previous studies, suggesting that the TNM stage-based clinical model is insufficient for the prediction of individual prognosis.

Most studies have indicated that biological characteristics of tumors can be revealed via quantitative medical image features, which will likely improve tumor
Figure 6 The performance of the nomogram. (A and B) Receiver operating characteristic (ROC) curves for the nomogram (A) and clinical model (B) show the predictive accuracy of each model in terms of the area under the curve (AUC) at predicting 3-year OS in the training and validation cohorts. (C and D) Calibration curves for the nomogram show the agreement between the estimated and observed 1-, 2-, 3-year outcomes in the training cohort (C) and the validation cohort (D). (E) Decision curve analysis for nomogram. The nomogram had a good net benefit compared with clinical model, deep learning signature and simple strategies such as follow-up of all patients (grey line) or no patients (horizontal black line) across the majority range of threshold probabilities.

prognostic prediction. Jong et al. reported that the hand-crafted signature showed prognostic value in 195 patients with lung adenocarcinoma (C-index, 0.576). We extracted deep learning features from CT images of 231 patients and selected robust features to build deep learning signature. The discrimination performance of the deep learning signature in our study (C-index, 0.748 and 0.695 for the training and external validation cohort) was better than the performance of hand-crafted signature in Jong’s study. Recently, Hosny et al. reported that CT-based deep learning signature may be used for risk stratification in NSCLC patients, and found that deep learning signature was significantly associated with binary 2-year OS (AUC=0.71). In our study, the predictive accuracy of deep learning signature (AUC, 0.759 and 0.785 for 3-year OS in the training and external validation cohort) was slightly higher than theirs. Meanwhile, the output of our finally constructed model (nomogram) was predicting the OS as a continuous variable, which was closer to the real clinical situation than binary analysis of 2-year OS. Moreover, deep learning-based nomogram built in our study showed a good estimation power and good discrimination performance, making it to be a practical and user-friendly tool for clinicians. The clinical usefulness of the nomogram for OS prediction was further confirmed by decision curve analysis.

In the present study, the deep learning signature could be utilized for the mortality risk-stratification of resected NSCLC patients within subgroups as defined by clinical risk factors. The important finding of our study is that the
deep learning signature identified low-risk patients with better OS in stage I–III NSCLC, which reflected that heterogeneity of survival outcome presented in the same stage. Huang et al. found that a hand-crafted feature signature has the potential to be a biomarker of risk stratification for disease-free survival (DFS) in early-stage NSCLC patients. However, their hand-crafted feature signature did not successfully stratify patients with stage II NSCLC.

Some limitations existed in this study. Firstly, the sample size was relatively small. Thus, these findings need further multi-institutional validation with a larger sample size. Secondly, the reconstructed image thickness was different between the two participant hospitals and even in the same hospital. However, the difference was not uncommon in current clinical practice, and we have drawn meaningful conclusions even under this imperfect condition. Thirdly, the deep learning algorithms often find their own rules and do not leave an audit trail to explain the decisions (namely, the black box problem), which is inherently opaque and has not yet been overcome. Finally, we used 2D features rather than 3D features in this study. Although 3D features taking the whole tumor into consideration may be more informative, several previous studies have shown that 2D features could also provide significant prognostic information compared to 3D features.

Conclusion
The deep learning signature could be used for risk stratification in resected NSCLC patients. Moreover, the nomogram combining the deep learning signature with clinical risk factors (TNM stage, lymphatic vessel invasion and differentiation grade) can be used to assist clinical decision-making.

Acknowledgments
This work was supported by National Natural Scientific Foundation of China [82072090 and 81601469], Natural Science Foundation of Guangdong Province in China [2018A030313511], Guangzhou Science and Technology Project of Health [2019A011002], Clinical Research Startup Program of Southern Medical University by High-level University Construction Funding of Guangdong Provincial Department of Education [LC2016PY034].

Disclosure
The authors report no conflicts of interest in this work.

References