Standardization of imaging features for radiomics analysis

Akihiro Haga1, PhD, Wataru Takahashi2, MD, PhD, Shuri Aoki3, MD, Kanabu Nawa1, PhD, Hideomi Yamashita1, MD, PhD, Osamu Abe4, MD, PhD, and Keiichi Nakagawa2, MD, PhD

1Graduate School of Biomedical Science, Tokushima University, Tokushima, Japan; 2Department of Radiology, The University of Tokyo Hospital, Tokyo, Japan

Abstract: Radiomics has the potential to provide tumor characteristics with noninvasive and repeatable way. The purpose of this paper is to evaluate the standardization effect of imaging features for radiomics analysis. For this purpose, we prepared two CT databases; one includes 40 non-small cell lung cancer (NSCLC) patients for whom tumor biopsies was performed before stereotactic body radiation therapy in The University of Tokyo Hospital, and the other includes 29 early-stage NSCLC datasets from the Cancer Imaging Archive. The former was used as the training data, whereas the later was used as the test data in the evaluation of the prediction model. In total, 476 imaging features were extracted from each data. Then, both training and test data were standardized as the min-max normalization, the z-score normalization, and the whitening from the principle component analysis. All of standardization strategies improved the accuracy for the histology prediction. The area under the receiver observed characteristics curve was 0.725, 0.789, and 0.785 in above standardizations, respectively.

Keywords: Radiomics, Quantitative imaging, Standardization, Histology prediction, Machine learning

INTRODUCTION

Radiomics can provide characteristics of entire tumors and of spatial and temporal intratumoral heterogeneity with noninvasive and repeatable way (1, 2). Radiomics converts medical imaging data into a high-dimensional feature space (typically more than two hundreds) using a large number of automatically extracted data characterization algorithms. Extracted radiomics features may be related with the outcome of tumor phenotype (3), treatment response (4), differentiate benign and malignant tumors (5), and histology (6). Radiomics have drawn an interest due to their possibility of uncovering tumor characteristics that may have otherwise failed to be appreciated by the naked eye.

On the other hand, there are several problems to be overcome in order to improve the radiomics prognosis of the outcome (7-9). One is that the radiomics signature has been sensitive to the delineation of the volume of interest, which is commonly subject to interobserver delineation variability. Second is a variation in medical images used in the radiomics analysis, that is, regarding image quantification or normalization. Third is a limited accessibility to the medical database. Above problems relate each other having inherent difficulties.

Previous study (10) showed the application of radiomics for predicting the histology of early-stage non-small cell lung cancer (NSCLC) by analyzing the computed tomography (CT) images with interobserver variability for tumor delineation. Its finding showed that inter-observer variability in delineation is a significant factor affecting the radiomics performance. However, one of the limitations of above study was the small cohort size.

Therefore, this study shows the validation results by using the different database: The Cancer Imaging Archive (11). It is critical to develop standardized method of the imaging features, and the quantitative imaging technique is essential in further development of radiomics with the multi-database analysis.

The purpose of this paper is to evaluate the effect of standardization of imaging features for radiomics analysis.

MATERIALS AND METHODS

Databases

Figure 1 denotes the schematic workflow of the present study, which employed the different two CT databases: one is the database from the University of Tokyo Hospital including 40 lung CT images acquired for the treatment planning in stereotactic body radiation therapy (adenocarcinoma: 21, squamous cell carcinoma: 19). The other one is the database form TCIA including 29 lung CT images (adenocarcinoma: 15, squamous cell carcinoma: 14, from NSCLC-Radiomics). Both databases were restricted to include early-stage NSCLC and contouring-completed patient datasets. Especially, we note that the database from the University of Tokyo Hospital includes four segmentations for a tumor in order to evaluate the interobserver variability: patient cohort and the imaging condition can be found in the references (1, 10). The study was ethically approved by the institutional review board at the University of Tokyo Hospital (reference number 3372).

Extraction and selection of the imaging features

In the previous study, we established the extraction and selection strategy in CT imaging features for lung cancer histology analysis (10). First, we perform a three-dimensional wavelet transform in each CT. With these decomposed images and the corresponding original image, second, 476 features including size-shape, histogram base, and texture base features were extracted from...
each dataset. Third, the features are selected by a univariate analysis with a certain threshold (a p-value < 0.1 in the random permutation test), interobserver delineation variability (to be selected by, at least, three out of four segmentations), and correlation analysis (a cross correlation value < 0.7). The selected features are seen in the reference (10), which were selected with The University of Tokyo Hospital database, and are used in the present investigation for the standardization of the imaging features.

**Standardization of the imaging feature**

The features selected above may be dependent on the imaging condition (an X-ray energy, a tube current, a reconstruction filter, etc.). Normalization of the features need to be applied, if the different protocols were used in imaging. It was normalized by the following procedure: The University of Tokyo database as a training database and the TCIA database as a test database, which have a different protocol each other, were employed. Both training and test data were standardized as the max-min normalization, where each feature was normalized as the range from 0 to 1, z-score normalization, where each feature was normalized as \( z = \frac{x - \mu}{\sigma} \), where \( x \), \( \mu \), and \( \sigma \) are the feature, the mean, and the standard deviation, respectively, and the whitening from the principle component analysis (PCA). Although the normalization should be performed in each database with same protocol, from the viewpoint of the data mixing based on the variance maximization, the whitening was performed by joining both databases.

**Evaluation**

In the present study, the “CORElearn” package in R was used. With this training data, a histology prediction model was constructed using the machine learning (the random forest model with “GainRatio”, “DkM”, “Gini”, “MDL”, “Relief”, “ReliefFmerit”, “ReliefFdistance”, “ReliefFexpRank” selections). The prediction performance was evaluated by the area under the receiver observed characteristics curve (AUC), accuracy, sensitivity, and specificity in the test cohort.

**RESULTS**

Table I shows the result of the histology classification with and without standardization of the imaging features, where the average AUC, accuracy, sensitivity, and specificity values of 8 decision-tree evaluation indices (“GainRatio”, etc.) are indicated. The model was optimized with The University of Tokyo Hospital database and the evaluation was performed with the TCIA database. Although the prediction result was acceptably high without standardization (AUC > 0.71), it was further improved by all of the normalization methods. Among them, the z-score normalization gave the best prediction and the data whitening gave the comparable result with the z-score normalization in AUC.

**DISCUSSION**

In radiomics, the first step is to derive a large amount of imaging features from medical images. Because all of the features are not always effective, a feature filtration process is needed in the following stage, where robust and few-redundant quantitative features should be selected. Actually, the imaging protocols are various, and as a result, the selected features might be database-dependent.

This study focused on the effectiveness of a standardization in imaging feature spaces for the radiomics analysis by using different two CT databases. The result suggests that the standardization of imaging features is influential in the histology prediction from CT images, and the data whitening as well as the z-score normalization is one of the strong ways in multi-database analysis. The z-score normalization which standardizes each feature independently may be ineffective when the features correlate each other.

On the other hand, using PCA we can make a more substantial normalization of the features to give it zero mean and unit covariance, so that different transformed features become decorrelated (12). The reason why the z-score normalization provided the comparable result with the whitening in the present study is thus, found to be reasonable because the correlation analysis was inserted in the feature selection process. This fact does not mean that the present result is trivial. We emphasize that such the reasonable result strongly supports our feature selection process surely selected the appropriate features for the histology prediction.

In the present study, the standardization in feature space has been investigated. From the viewpoint of a quantitative imaging, on the other hand, more fundamental approach can be taken into account. For instance, the electron density transformation is more essential to standardize the CT imaging features, because unlike the CT value, the electron density is a quantitative value independent of the imaging protocols. Furthermore, the recent development of multi-X-ray energy CT can provide the effective atomic number reconstruction as well as the electron density, directly (13). Such the approaches would be necessary for the development
of the standardized imaging database.

**CONCLUSION**

Radiomics analysis has shown that robust features have a high prognostic power in predicting early-stage NSCLC histology subtypes. The performance was able to be improved by standardizing the data in the feature space.

**COMPETING INTERESTS**

The authors declare that they have no competing interests.

**REFERENCES**


<table>
<thead>
<tr>
<th></th>
<th>W/o standardization</th>
<th>Min-max</th>
<th>Z-score</th>
<th>Whitening</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AUC</strong></td>
<td>0.715</td>
<td>0.725</td>
<td>0.789</td>
<td>0.785</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td>0.547</td>
<td>0.681</td>
<td>0.781</td>
<td>0.664</td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td>0.208</td>
<td>0.917</td>
<td>0.817</td>
<td>0.475</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td>0.911</td>
<td>0.429</td>
<td>0.741</td>
<td>0.866</td>
</tr>
</tbody>
</table>

"W/o standardization", "Min-max", "Z-score", and "Whitening" means the result using the features without standardization, the min-max normalization, z-score normalization, and the whitening by the principle component analysis.