Fig. 1 : Schematic illustration for the evaluation of the standardization in radiomics feature.
The radiomics features are extracted from two CT image databases acquired by the different imaging protocol each other, and then, are standardized by the min-max normalization, z-score normalization, or the whitening (sphering). One database is used as the training data, whereas the other database is used as the test data in a machine learning model.

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 min-max normalization, where each feature was normalized as the range from 0 to 1, z-score normalization, where each feature was normalized as $z = (x - \bar{x})/s$, where x, $\bar{x}$, and s 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 effectivity 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

Table 1 : AUC, accuracy, sensitivity, and specificity obtained by the random forest model for the test data.

	W/o standardization	Min–max	Z–score	Whitening
AUC	0.715	0.725	0.789	0.785
Accuracy	0.547	0.681	0.781	0.664
Sensitivity	0.208	0.917	0.817	0.475
Specificity	0.911	0.429	0.741	0.866

"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.

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

1. Aerts HJ, *et al* : Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 5 : 4006, 2014
2. Lambin P, *et al* : Radiomics Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48 : 441-446, 2012
3. Ellingson BM, *et al* : Probabilistic radiographic atlas of glioblastoma phenotypes. *Am J Neuroradiol* 34 : 533-540, 2013
4. Boellaard R : Need for standardization of 18F-FDG PET/CT for treatment response assessments. *J Nucl Med* 52 Suppl 2 : 93S-100S, 2011
5. Xu R, *et al* : Texture analysis on 18F-FDG PET/CT images to differentiate malignant and benign bone and soft-tissue lesions. *Ann Nucl Med* 28 : 926-935, 2014
6. Wu W, *et al* : Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. *Front Oncol* 6 : 71, 2016
7. Yip SSF, Aerts HJWL : Applications and limitations of radiomics. *Phys Med Biol* 61 : R150-66, 2016
8. Kumar V, *et al* : Radiomics : The process and the challenges. *Magn Reson Imaging* 30 : 1234-1248, 2012
9. Traverso A, *et al* : Repeatability and Reproducibility of Radiomic Features : A Systematic Review. *Int J Radiat Oncol Biol Phys* 104 : 1143-1158, 2018
10. Haga A, *et al* : Classification of early stage non-small cell lung cancers on computed tomographic images into histological types using radiomic features : interobserver delineation variability analysis. *Radiol Phys Technol* 11 : 1-9, 2017
11. Clark K, *et al* : The Cancer Imaging Archive (TCIA) : Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging* 26 : 1045-1057, 2013
12. Sakata D, *et al* : Effective atomic number estimation using kV-MV dual-energy source in LINAC, *Physica Medica* 39 : 9-15, 2017
13. Bishop CM : *Pattern Recognition and Machine Learning. Pattern Recognition* 4 : Springer, 2006