



Deep Learning Algorithm for Reducing CT Slice Thickness: Effect on Reproducibility of Radiomic Features in Lung Cancer

Sohee Park, MD¹, Sang Min Lee, MD¹, Kyung-Hyun Do, MD¹, June-Goo Lee, PhD², Woong Bae, PhD³, Hyunho Park, MD³, Kyu-Hwan Jung, PhD³, Joon Beom Seo, MD¹

¹Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea;

²Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea; ³VUNO Inc., Seoul, Korea

Objective: To retrospectively assess the effect of CT slice thickness on the reproducibility of radiomic features (RFs) of lung cancer, and to investigate whether convolutional neural network (CNN)-based super-resolution (SR) algorithms can improve the reproducibility of RFs obtained from images with different slice thicknesses.

Materials and Methods: CT images with 1-, 3-, and 5-mm slice thicknesses obtained from 100 pathologically proven lung cancers between July 2017 and December 2017 were evaluated. CNN-based SR algorithms using residual learning were developed to convert thick-slice images into 1-mm slices. Lung cancers were semi-automatically segmented and a total of 702 RFs (tumor intensity, texture, and wavelet features) were extracted from 1-, 3-, and 5-mm slices, as well as the 1-mm slices generated from the 3- and 5-mm images. The stabilities of the RFs were evaluated using concordance correlation coefficients (CCCs).

Results: The mean CCCs for the comparisons of original 1 mm vs. 3 mm, 1 mm vs. 5 mm, and 3 mm vs. 5 mm images were 0.41, 0.27, and 0.65, respectively ($p < 0.001$ for all comparisons). Tumor intensity features showed the best reproducibility while wavelets showed the lowest reproducibility. The majority of RFs failed to achieve reproducibility ($CCC \geq 0.85$; 3.6%, 1.0%, and 21.5%, respectively). After applying the CNN-based SR algorithms, the reproducibility significantly improved in all three pairings (mean CCCs: 0.58, 0.45, and 0.72; $p < 0.001$ for all comparisons). The reproducible RFs also increased (36.3%, 17.4%, and 36.9%, respectively).

Conclusion: The reproducibility of RFs in lung cancer is significantly influenced by CT slice thickness, which can be improved by the CNN-based SR algorithms.

Keywords: *Computed tomography; Radiomics; Slice thickness; Deep learning*

INTRODUCTION

Radiomics is an evolving field that involves the use of high-throughput computing to extract a large number of quantitative features from medical imaging data (1, 2). Recent studies have shown that radiomic features

(RFs) can serve as prognostic factors in lung cancer and nasopharyngeal cancer (2-4). In the study by Huang et al. (4), the radiomics signature enhanced performance in predicting disease-free survival when incorporated into a nomogram with clinicopathologic factors. However, these studies analyzed CT images acquired with a wide range of

Received March 20, 2019; accepted after revision July 12, 2019.

This study received funding from the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (Grant number: NRF-2016R1A2B1016355) and the Korea Health technology R&D Project, Ministry for Health & Welfare Affairs, Republic of Korea (Grant number: HI18C0673).

Corresponding author: Sang Min Lee, MD, Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea.

• Tel: (822) 3010-4382 • Fax: (822) 3010-6645 • E-mail: sangmin.lee.md@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

parameter settings.

Since RFs seem to show poor reproducibility across different CT protocols (5), there is concern over the generalizability of radiomics results and their application across multiple centers. In the phantom-based study conducted by Berenguer et al. (5), only 71 of 177 RFs were reproducible when extracted from CT images acquired using different parameters, including 2-, 3-, 5-, and 8-mm slice thicknesses. Although the results of phantom studies cannot necessarily be directly translated to the clinical setting, this study demonstrated the need for standardization of radiomics methodology. Such standardization may improve the reproducibility and diagnostic performance of RFs.

The acquisition thickness of CT images is one aspect affecting reproducibility. He et al. (6) reported that thin-slice (1.25 mm) CT-based radiomic signatures showed significantly better diagnostic performance than thick-slice CT (5 mm) signatures in the discrimination of benign and malignant solitary pulmonary nodules. These results imply that thin-slice CT contains more information for radiomics analysis; however, the acquisition of thin-slice CT images involves greater storage requirements for collection and management of the data.

In an effort to reduce the variability of RFs resulting from differences in slice thickness and to obtain more information from CT images, we considered whether thick-slice images (3 or 5 mm) could be converted to 1-mm slice-thickness images using super-resolution (SR) reconstruction, and whether this would reduce the differences in RFs associated with different CT slice thicknesses (7). SR reconstruction is a post-processing technique that allows a high-resolution image to be produced from a sequence of low-resolution images, and its implementation with convolutional neural network (CNN) algorithms has been investigated in several medical imaging studies (8-11).

Therefore, the aim of our study was to retrospectively assess the effect of CT slice thickness on the reproducibility of RFs of lung cancer, and to investigate whether CNN-based SR algorithms can improve the reproducibility of RFs obtained from images with different CT slice thicknesses.

MATERIALS AND METHODS

This retrospective study was approved by the Institutional Review Board of our center, which waived the requirement for patients' informed consent (approval number: 2018-1283).

Study Population

A retrospective search of the electronic medical records of our hospital for the period between July 2017 and December 2017 was performed by one radiologist (who had 2 years of experience in chest CT). A total of 445 patients who had undergone preoperative chest CT before surgical resection of primary lung cancer were identified.

The study population was enrolled on the basis of the following criteria: 1) identical CT parameter settings; contrast-enhanced CT images reconstructed with a B50f kernel and slice thicknesses of 1, 3, and 5 mm; and 2) lung cancers measuring up to 3 cm in the greatest diameter on CT. Exclusion criteria for this analysis were endobronchial or cavitory lesions, poor-quality CT images, and lesions that were difficult to clearly delineate because of accompanying lung parenchymal collapse or obstructive changes.

Among the 319 lung cancers identified as meeting the study criteria, 148 were solid and 171 were ground-glass nodules (GGNs), as classified by two radiologists in consensus. Subsequently, 50 nodules were randomly selected from each of the solid and GGN groups by using a pseudo-random number generating algorithm (Fig. 1). The final study population of 100 patients (mean age, 62.2 ± 10.6 years; range, 29-77 years) was enrolled (Table 1).

CNN-Based SR Technique

For development and validation of the model, CT data were collected from 100 patients without lung cancers. These 100 CT datasets were evenly divided into four categories according to the radiation dose (standard or low-dose) and the use of contrast enhancement (contrast-enhanced or not). Eighty-eight patients were randomly selected (22 from each category) to train the CNN model, with the remaining 12 patients (three from each category) being used to validate the model performance.

The SR algorithm for CT slice thickness is aimed at improving the resolution only in the depth direction, which can be interpreted as spatial information in the coronal and sagittal planes. The SR network is divided into preprocessing, non-linear mapping, and reconstruction parts (Fig. 2). The preprocessing part deals with variance from input images with different scales, and the non-linear mapping and reconstruction parts deal with the differences between the input images at low resolutions (3 or 5 mm) and the target image (1 mm). Further details are provided in the Supplementary Materials and Supplementary Table 1. To assess whether the CNN-based SR algorithm is better than

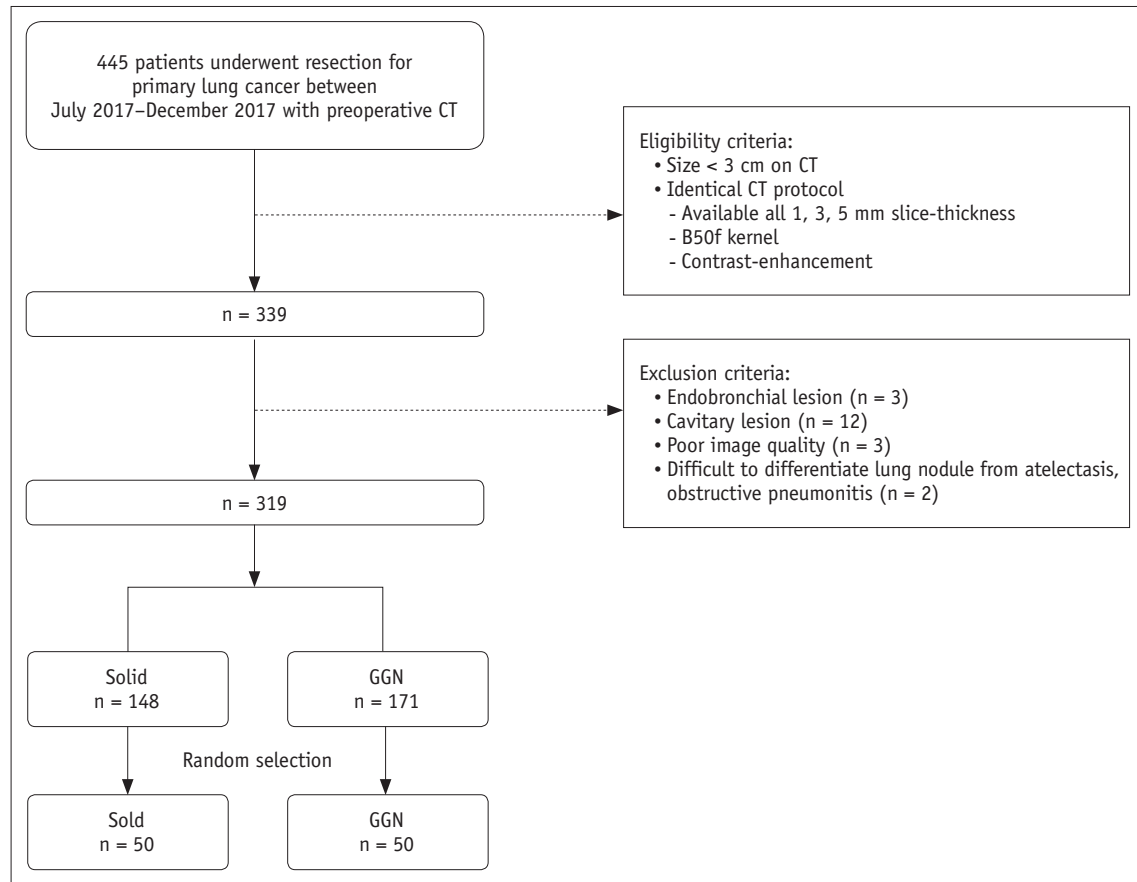
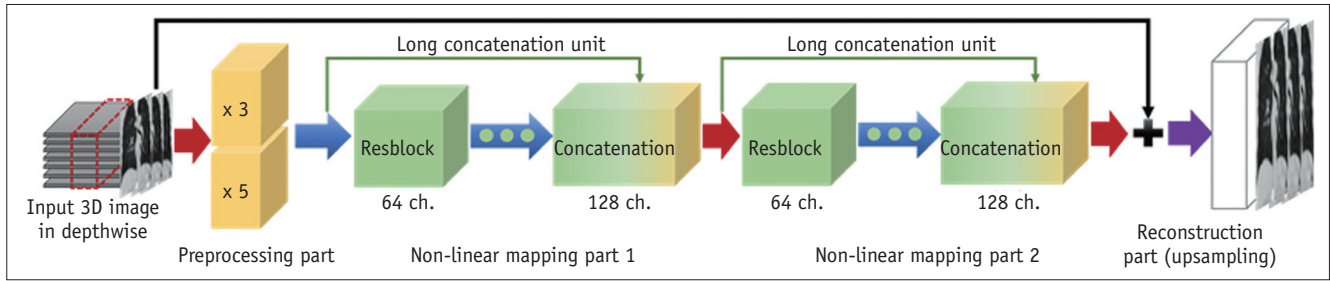


Fig. 1. Flow diagram for patient inclusion. GGN = ground-glass nodule

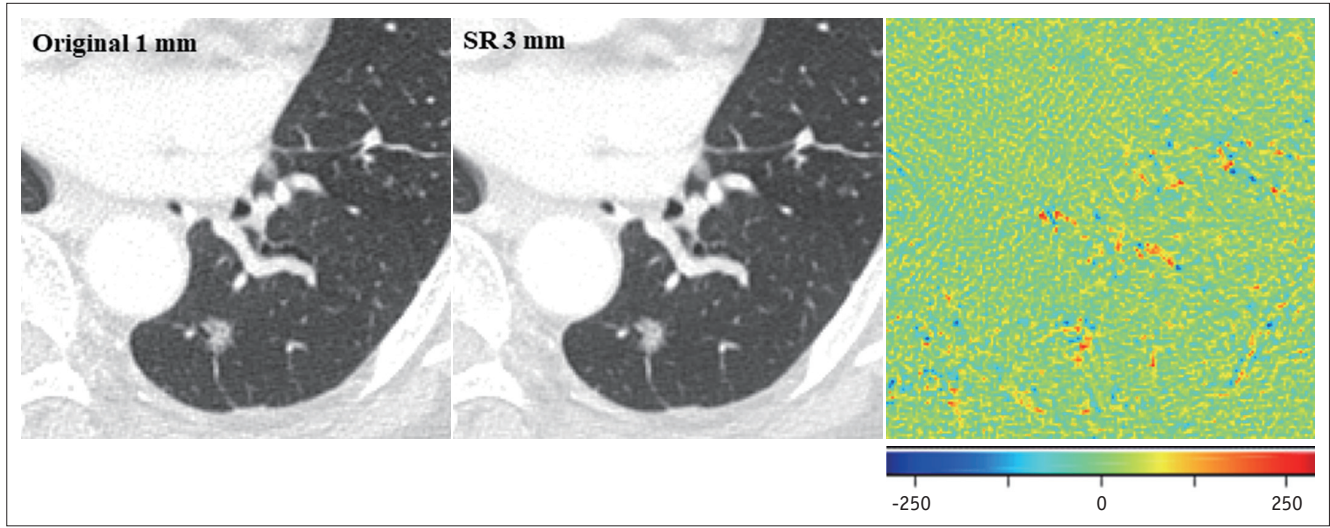
Table 1. Baseline Characteristics of Study Population (n = 100)

Characteristic	GGN (n = 50)	Solid (n = 50)	P
Age, years, mean ± SD	60.4 ± 10.4	64.0 ± 10.6	0.092
Sex (%)			0.007
Male	11 (22.0)	24 (48.0)	
Female	39 (78.0)	26 (52.0)	
Tumor size, mm, mean ± SD	16.3 ± 5.3	20.7 ± 6.9	0.001
Pathology (%)			
Minimally invasive adenocarcinoma	7 (14.0)		
Invasive adenocarcinoma	40 (80.0)	37 (74.0)	
Mucinous invasive adenocarcinoma	3 (6.0)	6 (12.0)	
Squamous cell carcinoma		4 (8.0)	
Small cell lung cancer		2 (4.0)	
Pleomorphic carcinoma		1 (2.0)	
Location (%)			
Right upper lobe	14 (28.0)	12 (24.0)	
Right middle lobe	1 (2.0)	6 (12.0)	
Right lower lobe	13 (26.0)	15 (30.0)	
Left upper lobe	14 (28.0)	12 (24.0)	
Left lower lobe	8 (16.0)	5 (10.0)	

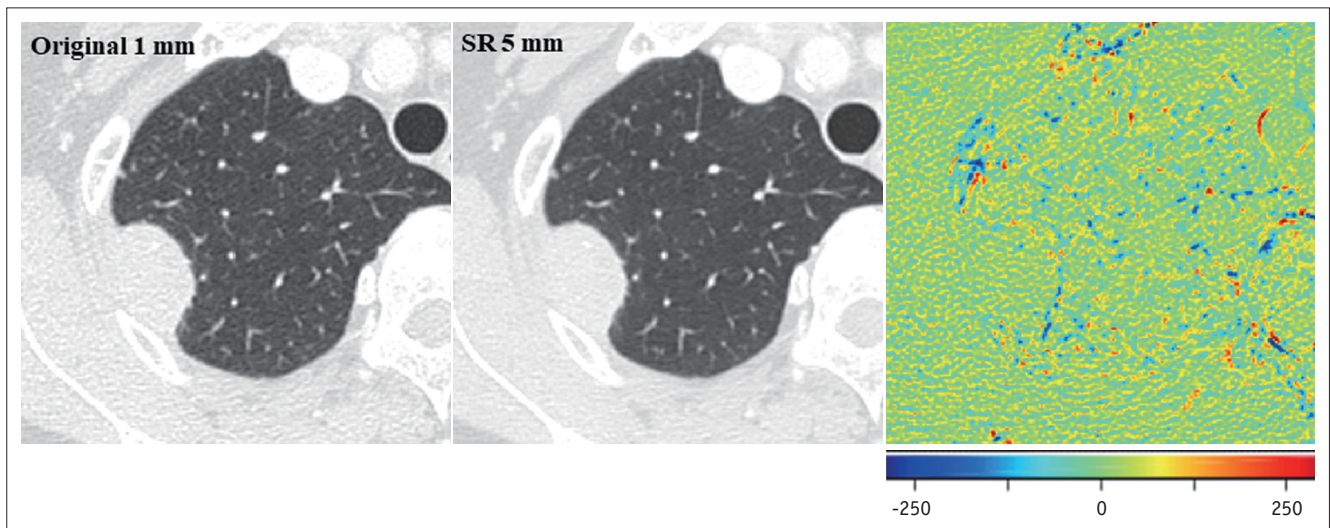
GGN = ground-glass nodule, SD = standard deviation



A



B



C

Fig. 2. Development of CNN algorithm for image conversion.

A. Network consists of preprocessing, non-linear mapping, and reconstruction parts. Preprocessing part deals with variance from input images with different scales, and non-linear mapping and reconstruction parts deal with differences between input images at low resolution (3 or 5 mm) and target image (1 mm). B. Ground-truth image, converted image, and difference map are presented. In difference map, green indicates zero, blue represents negative values, and red indicates positive values. When comparing original 1-mm and SR 3-mm images, most of lung parenchyma and small vascular structures show near-zero difference. Regions showing residual differences are present in pulmonary nodule, large vascular structures, fissure, and border between lung parenchyma and pleura (root-mean-square error: 30.83 HU). C. Comparison between original 1-mm and SR 5-mm images shows slightly more prominent differences in background lung parenchyma, including vascular structures, with speckled error regions remaining along margin of lung nodule (root-mean-square error: 37.65 HU). SR 3 mm refers to images converted from original 3-mm slice thickness, SR 5 mm refers to images converted from original 5-mm slice thickness. CNN = convolutional neural network, SR = super-resolution, 3D = three-dimensional

the conventional simple interpolation algorithm, a trilinear interpolation algorithm was applied to convert slice thickness.

CT Acquisition

Chest CT images were obtained with SOMATOM Definition AS and Sensation-16 scanners (Siemens Healthineers, Forchheim, Germany), with settings of 120 kVp and 150–200 mAs, a pitch of 0.875–1, and collimation of 1–1.25 mm. Images were reconstructed using a B50f kernel with slice thicknesses of 1, 3, and 5 mm. All CT data were acquired in the supine position at full inspiration. Scan coverage was from the lung base to the level of the thoracic inlet. Intravenous contrast media (100 mL) was injected at a rate of 3 mL/sec, and scanning started after a delay of 50 seconds.

Lung Nodule Segmentation and Radiomics Analysis

Tumor segmentation was performed on 1-, 3-, and 5-mm images using commercial software (AVIEW, Coreline Soft, Seoul, Korea). For each tumor, the software automatically drew regions of interest covering the entire range of the tumor on the axial CT images. Then, one radiologist manually adjusted the boundary regions on each section (12, 13). Two radiologists confirmed the masks for the final regions of interest in consensus. A total of 718 RFs consisting of tumor shape, intensity, texture, and wavelet features were extracted automatically, as previously described (2).

Since variability in tumor segmentation can also induce differences in the RFs, the mask boundaries on the original 3- and 5-mm slice-thickness images were adjusted with reference to the masks of the original 1-mm slice-thickness images. For the same reason, the masks of the original 1-mm images were copied to the converted 1-mm images.

Statistical Analysis

Root-mean-square errors (RMSEs) were obtained as a quantitative measurement to assess differences between the original 1-mm and converted 1-mm images. The reproducibility of each individual RF was calculated using concordance correlation coefficients (CCCs) (14), with a CCC of ≥ 0.85 indicating a reproducible RF (15). The CCCs of RFs were compared using a paired *t* test or Wilcoxon test to assess significant differences in the reproducibility of RFs in all possible pairings from the original images (1 mm vs. 3 mm; 1 mm vs. 5 mm; and 3 mm vs. 5 mm).

Subsequently, the CCCs of RFs obtained before and after the image conversion were compared. The performance of the CNN-based SR algorithm and that of a simple trilinear interpolation algorithm were compared in terms of CCCs. RFs between nodule types were also compared using an independent *t* test or Mann-Whitney tests. Sixteen shape features were excluded from this analysis because they had identical shapes, as they were copied from the original 1-mm mask.

All statistical analyses were performed using MedCalc statistical software (version 18.2.1; MedCalc Software, Ostend, Belgium). Data are presented as mean \pm standard deviation. $P < 0.05$ was considered to indicate statistical significance.

RESULTS

Global Performance of the CNN-Based SR Algorithm

The RMSEs were 38.2 ± 3.8 HU between images with an original 1-mm slice thickness and those converted from an original 3-mm slice thickness, and 47.1 ± 4.0 HU between the original 1-mm slice-thickness images and those converted from an original 5-mm slice thickness (Fig. 2).

Effect of Slice Thickness on RFs

The mean CCCs for the shape features were more than 0.85 in all three pairings, which we regarded as an acceptable difference for tumor segmentation. The mean CCCs for all classes of features (except for the shape features) between 1- and 3-mm, 1- and 5-mm, and 3- and 5-mm images were 0.41 ± 0.23 , 0.27 ± 0.21 , and 0.65 ± 0.22 , respectively, with all three pairings showing significant differences ($p < 0.001$ for all comparisons). Tumor intensity features showed the best reproducibility, while the CCCs for the wavelet features were the lowest in all three pairings.

In terms of nodule type, GGNs showed better reproducibility than solid nodules in all RF classes and in all slice-thickness pairings ($p < 0.001$ for 1 mm vs. 3 mm and 1 mm vs. 5 mm, and $p = 0.002$ for 3 mm vs. 5 mm). Table 2 presents the CCCs for the RF classes according to slice thickness.

The majority of RFs failed to reach the cut-off value for reproducibility in the original images, with 25, 7, and 151 out of 702 RFs (3.6%, 1.0%, and 21.5%, respectively) being considered as reproducible in comparisons between 1- and 3-mm, 1- and 5-mm, and 3- and 5-mm slice-thickness images, respectively.

Table 2. CCCs between Different Slice Thicknesses before and after SR Application

RF Classes	1 mm vs. 3 mm			1 mm vs. 5 mm			3 mm vs. 5 mm		
	Original	SR	<i>P</i>	Original	SR	<i>P</i>	Original	SR	<i>P</i>
Total									
Shape	0.93 ± 0.06			0.86 ± 0.12			0.92 ± 0.09		
Tumor intensity	0.75 ± 0.17	0.97 ± 0.05	< 0.001	0.60 ± 0.21	0.92 ± 0.11	< 0.001	0.90 ± 0.06	0.96 ± 0.05	< 0.001
Texture	0.62 ± 0.19	0.84 ± 0.15	< 0.001*	0.42 ± 0.18	0.72 ± 0.19	< 0.001	0.76 ± 0.16	0.88 ± 0.13	< 0.001*
Wavelet	0.38 ± 0.22	0.54 ± 0.37	< 0.001*	0.25 ± 0.19	0.41 ± 0.34	< 0.001	0.63 ± 0.22	0.71 ± 0.21	< 0.001*
Tumor intensity + texture + wavelet	0.41 ± 0.23	0.58 ± 0.37	< 0.001*	0.27 ± 0.21	0.45 ± 0.34	< 0.001*	0.65 ± 0.22	0.72 ± 0.21	< 0.001*
GGN									
Shape	0.91 ± 0.08			0.83 ± 0.15			0.88 ± 0.15		
Tumor intensity	0.79 ± 0.17	0.97 ± 0.05	< 0.001*	0.68 ± 0.19	0.89 ± 0.11	< 0.001*	0.91 ± 0.17	0.93 ± 0.05	0.303*
Texture	0.63 ± 0.23	0.80 ± 0.20	< 0.001	0.44 ± 0.22	0.66 ± 0.21	< 0.001*	0.74 ± 0.19	0.82 ± 0.17	0.001*
Wavelet	0.39 ± 0.23	0.54 ± 0.38	< 0.001*	0.26 ± 0.21	0.40 ± 0.33	< 0.001*	0.63 ± 0.22	0.67 ± 0.23	< 0.001*
Tumor intensity + texture + wavelet	0.42 ± 0.25	0.56 ± 0.37	< 0.001*	0.29 ± 0.22	0.44 ± 0.34	< 0.001*	0.65 ± 0.22	0.69 ± 0.23	< 0.001*
Solid									
Shape	0.94 ± 0.05			0.88 ± 0.10			0.95 ± 0.05		
Tumor intensity	0.58 ± 0.19	0.92 ± 0.19	< 0.001	0.40 ± 0.19	0.87 ± 0.22	< 0.001	0.83 ± 0.06	0.95 ± 0.09	< 0.001
Texture	0.51 ± 0.19	0.82 ± 0.15	< 0.001*	0.31 ± 0.16	0.68 ± 0.22	< 0.001	0.71 ± 0.14	0.87 ± 0.13	< 0.001
Wavelet	0.35 ± 0.23	0.53 ± 0.36	< 0.001*	0.22 ± 0.20	0.41 ± 0.34	< 0.001*	0.59 ± 0.26	0.71 ± 0.23	< 0.001*
Tumor intensity + texture + wavelet	0.37 ± 0.23	0.57 ± 0.36	< 0.001*	0.24 ± 0.20	0.44 ± 0.34	< 0.001*	0.60 ± 0.25	0.73 ± 0.23	< 0.001*

**p* values are derived from paired *t* test or Wilcoxon test. CCC = concordance correlation coefficient, RF = radiomic feature, SR = super-resolution

Reproducibility of RFs after Image Conversion Using the CNN-Based SR Algorithm

Image conversion from slice thicknesses of 3 and 5 mm to 1 mm significantly improved reproducibility in all three pairings when all classes of features were pooled (mean CCCs increased from 0.27–0.65 to 0.45–0.72; *p* < 0.001 for all comparisons). This improvement was also observed in the subgroupings based on the classes of RFs and nodule types (Table 2). The CCCs between 1- and 3-mm slice-thickness images improved from 0.75 ± 0.17 to 0.97 ± 0.05, 0.62 ± 0.19 to 0.84 ± 0.15, and 0.38 ± 0.22 to 0.54 ± 0.37 for tumor intensity, texture, and wavelet features respectively (*p* < 0.001 for all comparisons). The CCCs between 1- and 5-mm slice-thickness images improved from 0.60 ± 0.21 to 0.92 ± 0.11, 0.42 ± 0.18 to 0.72 ± 0.19, and 0.25 ± 0.19 to 0.41 ± 0.34 for tumor intensity, texture, and wavelet features respectively (*p* < 0.001 for all comparisons).

The CCCs between 3- and 5-mm slice-thickness images also significantly increased, although the pre-conversion values were higher than the CCCs between 1- and 3-mm slice thicknesses (0.90 ± 0.06 to 0.96 ± 0.05, 0.76 ± 0.16

to 0.88 ± 0.13, and 0.63 ± 0.22 to 0.71 ± 0.21 for tumor intensity, texture, and wavelet features, respectively; *p* < 0.001 for all comparisons). Only the CCC for the tumor intensity features of GGNs was not significantly higher after image conversion (0.91 ± 0.17 to 0.93 ± 0.05, *p* = 0.303).

In accordance with this increase in the mean CCCs, the numbers of reproducible RFs also increased (Table 3, Fig. 3). Almost all tumor intensity features became reproducible (18, 16, and 18 out of 19 RFs for 1 mm vs. 3 mm, 1 mm vs. 5 mm, and 3 mm vs. 5 mm, respectively) after image conversion. In addition, the number of reproducible features in the texture class, which was particularly low in the comparisons between 1-mm vs. 3-mm and 1-mm vs. 5-mm slice-thickness images, significantly increased after image conversion (3 to 40, 0 to 17, and 23 to 45 out of 59 RFs for 1 mm vs. 3 mm, 1 mm vs. 5 mm and 3 mm vs. 5 mm, respectively).

For the wavelet features, both the mean CCC and the effect of image conversion varied among the subclasses of wavelet features, depending on the combination of high- or low-frequency components (Supplementary Table 2). The mean CCC tended to be better when more low-pass filters

Reducing Effect of CT Slice-Thickness in Radiomics

were applied.

When all features were pooled, the total number of reproducible features markedly increased from 16 to 197, 5

to 89, and 114 to 196 out of 624 RFs for 1 mm vs. 3 mm, 1 mm vs. 5 mm, and 3 mm vs. 5 mm, respectively.

Using simple interpolation, the overall reproducibility

Table 3. Numbers of RFs Meeting Reproducibility Criteria according to Different CCC Thresholds (Cut-Offs; 0.80, 0.85, and 0.90)

RF Classes	1 mm vs. 3 mm						1 mm vs. 5 mm						3 mm vs. 5 mm					
	Original			SR			Original			SR			Original			SR		
	0.80	0.85	0.90	0.80	0.85	0.90	0.80	0.85	0.90	0.80	0.85	0.90	0.80	0.85	0.90	0.80	0.85	0.90
Total																		
Tumor intensity (n = 19)	7 (36.8)	6 (31.2)	5 (26.3)	19 (100)	18 (94.7)	17 (89.5)	5 (26.3)	2 (10.5)	1 (5.3)	16 (84.2)	16 (84.2)	15 (78.9)	18 (94.7)	14 (73.7)	10 (52.6)	19 (100)	18 (94.7)	16 (84.2)
Texture (n = 59)	6 (10.2)	3 (5.1)	0 (0.0)	46 (78.0)	40 (67.8)	22 (37.3)	0 (0.0)	0 (0.0)	0 (0.0)	20 (33.9)	17 (28.8)	12 (20.3)	31 (52.5)	23 (39.0)	11 (18.6)	50 (84.7)	45 (76.3)	41 (69.5)
Wavelet (n = 624)	28 (4.5)	16 (2.6)	6 (1.0)	235 (37.7)	197 (31.6)	152 (24.4)	6 (1.0)	5 (0.8)	1 (0.2)	110 (17.6)	89 (14.3)	71 (11.4)	178 (28.5)	114 (18.3)	64 (10.3)	259 (41.5)	196 (31.4)	135 (21.6)
GGN																		
Tumor intensity (n = 19)	14 (73.7)	12 (63.2)	4 (21.1)	19 (100)	18 (94.7)	17 (89.5)	4 (21.1)	1 (5.3)	1 (5.3)	15 (78.9)	15 (78.9)	10 (52.6)	19 (100)	17 (89.5)	11 (57.9)	19 (100)	18 (94.7)	10 (52.6)
Texture (n = 59)	19 (32.2)	13 (22.0)	3 (5.1)	43 (72.9)	36 (61.0)	19 (32.2)	2 (3.4)	2 (3.4)	0 (0.0)	18 (30.5)	14 (23.7)	4 (6.8)	27 (45.8)	22 (37.3)	8 (13.6)	45 (76.3)	40 (67.8)	20 (33.9)
Wavelet (n = 624)	40 (6.4)	24 (3.8)	5 (0.8)	244 (39.1)	211 (33.8)	168 (26.9)	9 (1.4)	3 (0.5)	3 (0.5)	98 (15.7)	85 (13.6)	44 (7.1)	162 (26.0)	112 (17.9)	65 (10.4)	229 (36.7)	174 (27.9)	93 (14.9)
Solid																		
Tumor intensity (n = 19)	1 (5.3)	1 (5.3)	1 (5.3)	17 (89.5)	16 (84.2)	15 (78.9)	1 (5.3)	1 (5.3)	1 (5.3)	15 (78.9)	15 (78.9)	13 (68.4)	10 (52.6)	8 (42.1)	1 (5.3)	17 (89.5)	16 (84.2)	16 (84.2)
Texture (n = 59)	2 (3.4)	0 (0.0)	0 (0.0)	42 (71.2)	34 (57.6)	18 (30.5)	0 (0.0)	0 (0.0)	0 (0.0)	19 (32.2)	14 (23.7)	10 (16.9)	17 (28.8)	11 (18.6)	3 (5.1)	42 (71.2)	39 (66.1)	37 (62.7)
Wavelet (n = 624)	28 (4.5)	15 (2.4)	5 (0.8)	206 (33.0)	183 (29.3)	143 (22.9)	8 (1.3)	6 (1.0)	2 (0.3)	117 (18.8)	83 (13.3)	65 (10.4)	151 (24.2)	98 (15.7)	54 (8.7)	260 (41.7)	215 (34.5)	147 (23.5)

Data in parentheses are percentages.

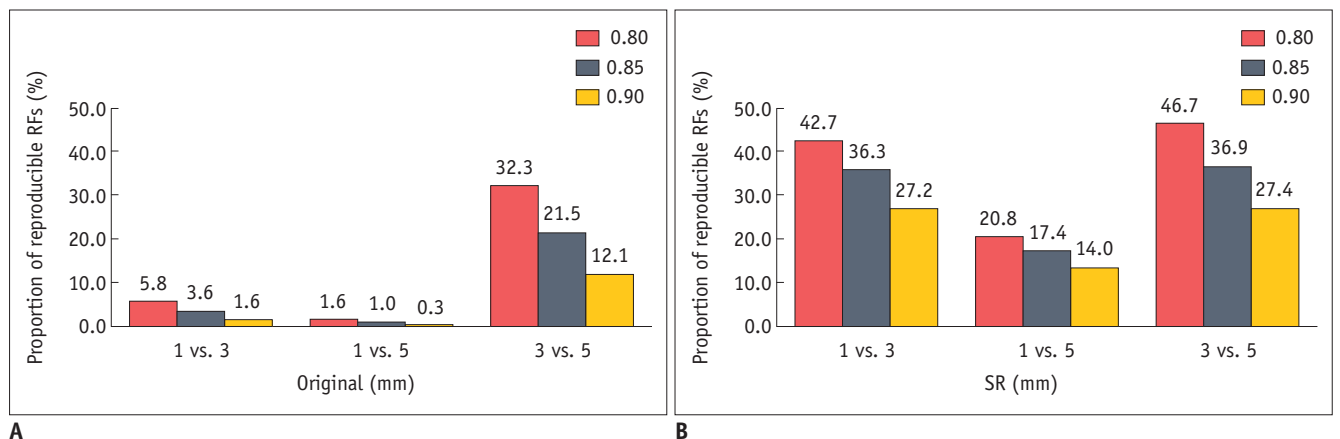


Fig. 3. Proportion of reproducible RFs (CCC ≥ 0.85) in different slice-thickness pairings before and after application of image conversion.

A. Majority of RFs failed to reach cut-off for reproducibility (CCC ≥ 0.85) in original images. After image conversion (**B**), proportion of reproducible RFs markedly increased in all three slice-thickness pairings. Similar tendency was observed using alternative cut-off values of 0.80 and 0.90. CCC = concordance correlation coefficients, RF = radiomic feature, SR = super-resolution

of the RFs was significantly lower than that with the SR algorithm ($p < 0.001$ for all comparisons) and was partly worse than those in all three pairings of different original slice-thickness images (Supplementary Table 3).

DISCUSSION

In this study, we demonstrated that the variability in RFs between different CT slice thicknesses was significant in lung cancer, and this variability can be reduced by the use of our CNN-based SR algorithm.

Previous studies showed that a large proportion of RFs (106/177) were prone to change according to the acquisition parameters (5), and that the average number of stable RFs was only 53.4 out of 114 (46.8%) when 1.5- and 3-mm slice thicknesses were compared, even when all sequences were acquired on a single CT scanner (16). These results are in line with our study, which indicates that the overall reproducibility of RFs between different slice-thickness images is not satisfactory (mean CCCs: 0.27–0.65). However, Lu et al. (17) reported that the mean CCCs of RFs were high (0.725–0.875) between pairings of different slice thicknesses (1.25, 2.5, and 5 mm) in 32 lung cancer patients. These discrepancies between our study and the study by Lu et al. (17) could mainly lie in differences in the numbers and composition of RFs, the status of contrast enhancement, and the CT vendor. We included a large number of wavelet features (624/702), which produced the lowest CCCs in the subgroup analysis, and used contrast-enhanced images taken on a scanner from a different vendor. Therefore, further studies with different acquisition settings are required to verify our results.

As expected, the agreements between 1- and 5-mm slice thicknesses were the lowest, while the 3- and 5-mm slice-thickness pairing showed the highest agreement. This agreement trend implies that the more two datasets overlap, the higher the similarity between the images. Tumor intensity and texture features yielded acceptable agreement levels between 3- and 5-mm slice thicknesses; however, in this pairing showing the best agreement, there was still a large proportion of non-reproducible RFs (78.5%, 551/702), and this implies that comparisons of radiomics results derived from CT images with different slice thicknesses may be unreliable.

To overcome this limitation, we developed the CNN-based SR algorithm to reduce the variability of the RFs. When the algorithm was designed, we chose 1 mm as a target

slice thickness because measurements of subcentimeter lung nodules should be obtained using thin-slice CT (≤ 1.5 mm), as recommended by the Fleischner Society (18), and thin-slice CT (1.25 mm) is more informative than 5-mm CT for discriminating between benign and malignant solitary pulmonary nodules (6).

Our CNN-based method was consistently effective in all subgroups from all three pairings of different slice thicknesses and showed a significant increase in reproducibility, except for the 3- and 5-mm slice-thickness comparison in the GGN group. Furthermore, the total number of reproducible RFs (CCC ≥ 0.85) was much higher in comparison with the original images, which provided only a small number of reproducible RFs across the various slice-thickness protocols. Therefore, the SR technique can be used to decrease the effect of slice thickness when RFs extracted from different slice-thickness images were compared with each other or analyzed together.

Another interesting finding of our study was that GGNs showed better reproducibility than solid nodules in all RF feature classes in all slice thicknesses, despite GGNs being vulnerable to slice thickness, as they appear faint and may be missed on thick-slice images. A possible explanation is that the volume-averaging effect changes the attenuation and its distribution of solid nodules more greatly than that of GGNs because the difference in attenuation between normal lung parenchyma and GGNs is smaller than that between normal lung parenchyma and solid nodules. However, other factors such as nodule shape or nodule size could influence the results. Matched case comparisons between solid nodules and GGNs may be necessary to confirm our results.

Our study has several limitations. First, our study included only contrast-enhanced images because preoperative CT for lung cancer was performed with contrast enhancement in routine practice. To compensate for this limitation, we used equal quantities of non-contrast and contrast-enhanced CTs for the training set. Thus, we expect the SR technique to work effectively on non-contrast images as well. A second limitation was the variability in tumor segmentation itself. The boundaries of a nodule on 1-, 3-, and 5-mm slice-thickness images are not inherently the same. However, given that substantial agreement in the shape features was achieved between all three pairings of original images (mean CCC ≥ 0.85 for all comparisons), the effects of the segmentation differences on the radiomics data were within acceptable ranges. Third, the shape features that were most sensitive

to slice thickness were excluded from the analysis, as the masks were drawn by referencing or copying the masks of the original 1-mm slice-thickness images, to ensure that the effect of slice thickness was evaluated independently of segmentation effects. When applying the SR technique in real practice, segmentation variability, as well as slice thickness, will affect the reproducibility of the extracted RFs. Therefore, in future studies, it will be necessary to investigate the effects of both of these through separate tumor segmentations. Fourth, the trilinear interpolation algorithm applied in our study is one of various interpolation methods and may not be the optimal method for slice-thickness reduction. Therefore, it might be hasty to mention that our CNN-based algorithm outperforms simple interpolation methods. Last, since our study is a feasibility study to investigate the effect of a CNN-based SR algorithm on the reproducibility of RFs, we did not assess whether the better reproducibility obtained with the SR technique will lead to improvements in the performance of radiomics-based tasks. Therefore, further studies are warranted to determine the clinical utility of our method.

In conclusion, the reproducibility of RFs in lung cancer is significantly influenced by CT slice thickness, which can be improved by the application of CNN-based SR algorithms.

Supplementary Materials

The Data Supplement is available with this article at <https://doi.org/10.3348/kjr.2019.0212>.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

ORCID iDs

Sang Min Lee

<https://orcid.org/0000-0001-7627-2000>

Sohee Park

<https://orcid.org/0000-0002-4009-1571>

Kyung-Hyun Do

<https://orcid.org/0000-0003-1922-4680>

June-Goo Lee

<https://orcid.org/0000-0002-1380-6682>

Woong Bae

<https://orcid.org/0000-0003-3128-9866>

Hyunho Park

<https://orcid.org/0000-0002-6395-478X>

Kyu-Hwan Jung

<https://orcid.org/0000-0002-6626-6800>

Joon Beom Seo

<https://orcid.org/0000-0003-0271-7884>

REFERENCES

- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;278:563-577
- Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006
- Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin Cancer Res* 2017;23:4259-4269
- Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology* 2016;281:947-957
- Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, Castro-García M, Villas MV, Mansilla Legorburo F, et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. *Radiology* 2018;288:407-415
- He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep* 2016;6:34921
- Kim J, Lee JK, Lee KM. *Accurate image super-resolution using very deep convolutional networks*. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR);2016 June 27-30:1646-1654;Las Vegas, NV, USA
- Bhavsar A, Wu G, Lian J, Shen D. Resolution enhancement of lung 4D-CT via group-sparsity. *Med Phys* 2013;40:121717
- Rueda A, Malpica N, Romero E. Single-image super-resolution of brain MR images using overcomplete dictionaries. *Med Image Anal* 2013;17:113-132
- Coupé P, Manjón JV, Chamberland M, Descoteaux M, Hiba B. Collaborative patch-based super-resolution for diffusion-weighted images. *Neuroimage* 2013;83:245-261
- Bahrami K, Shi F, Rekić I, Gao Y, Shen D. 7T-guided super-resolution of 3T MRI. *Med Phys* 2017;44:1661-1677
- Kim C, Lee SM, Choe J, Chae EJ, Do KH, Seo JB. Volume doubling time of lung cancer detected in idiopathic interstitial pneumonia: comparison with that in chronic obstructive pulmonary disease. *Eur Radiol* 2018;28:1402-1409
- Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* 2014;9:e102107
- Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255-268

15. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* 2016;6:23428
16. Larue RTHM, van Timmeren JE, de Jong EEC, Feliciani G, Leijenaar RTH, Schreurs WMJ, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncol* 2017;56:1544-1553
17. Lu L, Ehmke RC, Schwartz LH, Zhao B. Assessing agreement between radiomic features computed for multiple CT imaging settings. *PLoS One* 2016;11:e0166550
18. Bankier AA, MacMahon H, Goo JM, Rubin GD, Schaefer-Prokop CM, Naidich DP. Recommendations for measuring pulmonary nodules at CT: a statement from the Fleischner Society. *Radiology* 2017;285:584-600