

RESEARCH ARTICLE

A deep learning method for translating 3DCT to SPECT ventilation imaging: First comparison with $^{81\text{m}}\text{Kr}$ -gas SPECT ventilation imaging

Tomohiro Kajikawa^{1,2} | Noriyuki Kadoya² | Yosuke Maehara¹ | Hiroshi Miura³ |
Yoshiyuki Katsuta² | Shinsuke Nagasawa¹ | Gen Suzuki¹ | Hideya Yamazaki¹ |
Nagara Tamaki¹ | Kei Yamada¹

¹Department of Radiology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan

²Department of Radiation Oncology, Tohoku University Graduate School of Medicine, Sendai, Japan

³Department of Radiology, Japanese Red Cross Kyoto Daini Hospital, Kyoto, Japan

Correspondence

Tomohiro Kajikawa, Department of Radiology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, 465 Kajii-cho, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto 602-8566, Japan.
Email: tkjkw@koto.kpu-m.ac.jp

Funding information

Japan Society of for the Promotion of Science/Grant-in-Aid for Scientific Research (C), Grant/Award Number: 20K16733

Abstract

Purpose: This study aimed to evaluate the accuracy of deep learning (DL)-based computed tomography (CT) ventilation imaging (CTVI).

Methods: A total of 71 cases that underwent single-photon emission CT $^{81\text{m}}\text{Kr}$ -gas ventilation (SPECT V) and CT imaging were included. Sixty cases were assigned to the training and validation sets, and the remaining 11 cases were assigned to the test set. To directly transform three-dimensional (3D) CT (free-breathing CT) images to SPECT V images, a DL-based model was implemented based on the U-Net architecture. The input and output data were 3DCT- and SPECT V-masked, respectively, except for whole-lung volumes. These data were rearranged in voxel size, registered rigidly, cropped, and normalized in preprocessing. In addition to a standard estimation method (i.e., without dropout during the estimation process), a Monte Carlo dropout (MCD) method (i.e., with dropout during the estimation process) was used to calculate prediction uncertainty. To evaluate the two models' (CTVI_{MCD U-Net}, CTVI_{U-Net}) performance, we used fivefold cross-validation for the training and validation sets. To test the final model performances for both approaches, we applied the test set to each trained model and averaged the test prediction results from the five trained models to acquire the mean test result (bagging) for each approach. For the MCD method, the models were predicted repeatedly (sample size = 200), and the average and standard deviation (SD) maps were calculated in each voxel from the predicted results. The average maps were defined as test prediction results in each fold. As an evaluation index, the voxel-wise Spearman rank correlation coefficient (Spearman r_s) and Dice similarity coefficient (DSC) were calculated. The DSC was calculated for three functional regions (high, moderate, and low) separated by an almost equal volume. The coefficient of variation was defined as prediction uncertainty, and these average values were calculated within three functional regions. The Wilcoxon signed-rank test was used to test for a significant difference between the two DL-based approaches.

Results: The average indexes with one SD (1SD) between CTVI_{MCD U-Net} and SPECT V were 0.76 ± 0.06 , 0.69 ± 0.07 , 0.51 ± 0.06 , and 0.75 ± 0.04 for Spearman r_s , DSC_{high}, DSC_{moderate}, and DSC_{low}, respectively. The average

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.

indexes with 1SD between $CTVI_{U-Net}$ and SPECT V were 0.72 ± 0.05 , 0.66 ± 0.04 , 0.48 ± 0.04 , and 0.74 ± 0.06 for Spearman r_s , DSC_{high} , $DSC_{moderate}$, and DSC_{low} , respectively. These indexes between $CTVI_{MCD U-Net}$ and $CTVI_{U-Net}$ showed no significance difference (Spearman r_s , $p = 0.175$; DSC_{high} , $p = 0.123$; $DSC_{moderate}$, $p = 0.278$; DSC_{low} , $p = 0.520$). The average coefficient of variations with 1SD were 0.27 ± 0.00 , 0.27 ± 0.01 , and 0.36 ± 0.03 for the high-, moderate-, and low-functional regions, respectively, and the low-functional region showed a tendency to exhibit larger uncertainties than the others.

Conclusion: We evaluated DL-based framework for estimating lung-functional ventilation images only from CT images. The results indicated that the DL-based approach could potentially be used for lung-ventilation estimation.

KEYWORDS

deep learning, functional imaging, radiotherapy

1 | INTRODUCTION

Lung-functional imaging has been reported to provide beneficial information for radiotherapy (e.g., functional lung-avoidance radiotherapy treatment planning,^{1–3} and treatment response modeling⁴). Functional lung-avoidance radiotherapy, especially, is used to irradiate tumors to minimize dose deposition to high-functional lung regions by using guided functional imaging. The usefulness of this approach has been reported, and clinical practices and trials are underway.^{5,6}

Currently, most lung-ventilation and -perfusion imaging methods are based on the use of radioisotopes or contrast media, such as single-photon emission computed tomography (SPECT),⁷ positron emission tomography (PET),⁸ magnetic resonance imaging,⁹ and dual-energy CT.¹⁰ Although these imaging techniques provide much beneficial information for regional lung function, they have some disadvantages, including radiation exposure, high cost, relatively low spatial and/or temporal resolution, and the requirement of special equipment. Therefore, the routine use of these methods for radiotherapy is difficult, so a more convenient approach for lung-functional imaging is needed to perform functional lung-avoidance radiotherapy. Recently, to overcome these shortcomings, image-processing-based methods have been proposed to quantify the local lung-ventilation function from four-dimensional (4D) CT images.¹¹ This CT ventilation imaging (CTVI) technique is based on a deformable image registration algorithm that enables quantitation of the local lung-ventilation function from 4DCT images acquired for treatment planning. Using this technique, the special medication or implementations mentioned previously are unnecessary, so acquiring the lung-ventilation images becomes more convenient. Thus, 4DCT-based ventilation imaging has high accessibility and the potential to be used in radiation oncology departments because no additional imaging and financial costs for the treatment protocol are required. In addition, studies on lung-functional avoidance radiotherapy have reported the potential of substitutability for clinical ven-

tilation imaging, such as SPECT^{12,13} and PET,^{8,14} and this has been used in some prospective clinical trials (NCT02528942 and NCT02843568). However, 4DCT results in higher radiation exposure and imaging cost compared with three-dimensional (3D) CT.

In recent years, the use of deep learning (DL) for several tasks has become popular in radiotherapy, and particularly, convolutional neural networks (CNNs) have been primarily used for image processing.¹⁵ CNN consists of multilayered extraction and/or reconstruction features, and it is possible to perform end-to-end learning of features at various levels depending on the tasks. This technique has enabled CNN to provide comparable or superior performance to that of conventional approaches for various tasks, such as segmentation,¹⁶ dose prediction,^{17,18} dose calculation,¹⁹ and quality assurance.^{20,21} In addition, CNN was applied for lung-functional image generation and provided performance superior to that of conventional CTVI techniques. Zhong et al. transformed inhale and exhale CT images in 4DCT images to lung-ventilation images derived from CTVI directly.²² Liu et al. transformed 4DCT images to ventilation images derived from SPECT ventilation (V) directly.²³ On the other hand, Ren et al. proposed and investigated a method for generating SPECT perfusion images directly from CT images and showed that the method could generate perfusion images relatively accurately only with 3DCT images.^{24,25} Their results suggest that 3DCT images contain physiological and biological information necessary for estimating regional lung function.

Therefore, this study aimed to evaluate the accuracy of DL-based approach for directly transforming 3DCT images to krypton-81m gas (^{81m}Kr-gas) SPECT V images. The clinical SPECT V image was taken as true data, and the performance was evaluated qualitatively by visual inspection and quantitatively by indexes (Spearman rank's correlation coefficient and Dice similarity coefficient [DSC]). In addition, to interpret the prediction results, the prediction uncertainties were quantified and evaluated using the Monte Carlo dropout (MCD) approach.²⁶ To the best of our knowledge, this

TABLE 1 Clinical characteristics of the cases ($n = 71$): 67 patients had one SPECT/CT image set and 2 patients had 2 images

Characteristics	Median (range) or number (%)
Age (years)	72 (16–87)
Sex	Male: 18 (24.3%)/female: 53 (75.7%)
Smoking history	Yes: 20 (28.2%)/no: 50 (70.4%)/N/A: 1 (1.4%)
Diagnosis	Pulmonary hypertension: 30 (42.3%) Pulmonary embolism: 28 (39.4%) Systemic lupus erythematosus: 2 (2.8%) Others: 11 (15.5%)

Note: Others: paroxysmal supraventricular tachycardia, peripheral pulmonary stenosis, pulmonary arteriovenous fistula, atrial septal defect, severe tricuspid regurgitation, pulmonary vein stenosis, abdominal aortic aneurysm, transthyretin amyloid cardiomyopathy.

Abbreviations: CT, computed tomography; SPECT, single-photon emission computed tomography.

is the first study to translate SPECT V images directly from 3DCT images.

2 | MATERIALS AND METHODS

2.1 | Patients and imaging

This study was a retrospective analysis approved by our Institutional Review Board (ERB-C-1811-1). Sixty-nine patients who underwent lung $^{81\text{m}}\text{Kr}$ -gas ventilation SPECT/CT scans from 2016 to 2021 at our hospital were enrolled. These patients had all the required data (e.g., attenuation-corrected SPECT data and 3DCT images) for analysis. In total, 67 patients had 1 SPECT/CT image and 2 patients had 2 images (initial and follow-up scan) acquired at a ≥ 6 -month interval from the initial SPECT/CT examination. The images of a total of 71 cases were used. Accordingly, 60 cases were assigned to the training and validation sets randomly, and the remaining 11 patients were assigned to the test set for dividing the initial and follow-up scan in the same dataset to prevent data leakage. Among the 71 cases, 30 had pulmonary hypertension, 28 had pulmonary embolism, and 2 had systemic lupus erythematosus. The remaining patients had different types of lung diseases, such as paroxysmal supraventricular tachycardia and peripheral pulmonary stenosis. The clinical characteristics of the patients are summarized in Table 1.

We acquired $^{81\text{m}}\text{Kr}$ -gas ventilation images using a Discovery 670 SPECT/CT scanner (GE Healthcare, Milwaukee, Wisconsin). All scans were performed during the inhalation of 185 MBq of $^{81\text{m}}\text{Kr}$ -gas. The scans were acquired across 360° in six steps to cover the whole-lung volume at a frame rate of 20 s/step. An extended low-energy general-purpose collimator was used, and attenuation correction was applied for ventilation scans. All SPECT images were collected as a $64 \times 64 \times 64$ matrix with a pixel spacing of $8.84 \times 8.84 \times 8.84 \text{ mm}^3$

using a 3D-ordered subsets expectation maximization algorithm. The mean ± 1 standard deviation (1SD) of the non-defect ventilation volume percentage segmented using the image-specific intensity threshold (set at 50% of the 90th percentile value within the whole lung)^{12,14} for all SPECT V images was $68.0\% \pm 11.1\%$.

All CT images were acquired on 64-multidetector row CT scanners (Optima CT 540 or BrightSpeed; GE Health care, Milwaukee, Wisconsin) that compose the SPECT/CT system. All CT images were acquired during free-breathing and collected as a 512×512 matrix with a pixel spacing of $0.977 \times 0.977 \text{ mm}^2$ in axial view with axial spacing of 1.25 mm and reconstructed using the same filtered back-projection kernel for soft tissue. The X-ray tube voltage and current were set to 75 kV and 120 mA, respectively. The radiation exposure of the CT technique with this scanner was 3.06 mGy in volume CT dose index.

2.2 | Image preprocessing

The pipeline of the image preprocessing for model training and evaluation consisted of the following processes: parenchyma segmentation, registration, pixel resampling and resizing, and rescaling or normalization. Lung parenchyma segments were automatically generated from the CT images using 3D slicer software (<https://www.slicer.org>; ver. 4.11) extension Chest Imaging Platform (<http://www.chestimagingplatform.org>) and modified manually if necessary (e.g., deleting obvious digestive gas). To minimize the influence of misalignment between the SPECT V and CT images, we registered the alignment rigidly on Elastix 5.0.1 using PyElastix 1.2 (<https://github.com/almarklein/pyelastix>) as the python wrapper. To reduce the computational cost in DL model training, the CT images and lung masks were down-sampled and cropped into $2.5 \times 2.5 \times 2.5\text{-mm}^3$ and $128 \times 128 \times 128$ -sized matrices; these parameters were selected to include the whole-lung volume in the input data for all patients. Then, the SPECT V images were up-sampled and cropped under the same conditions as described previously. Voxel intensity values in CT are absolute and present important information (e.g., a single threshold of -856 HU indicates emphysematous lesions²⁷). Therefore, to maintain the relationship along CT values, CT images were not standardized but rather rescaled linearly using the following scale to smoothly train the DL-based model: $-999 \text{ HU} = 0$ and $-250 \text{ HU} = 1$. Conversely, voxel intensity values in SPECT imaging are relative rather than absolute compared with those in CT imaging, and it is possible that they have outliers (i.e., too high or too low local values). Therefore, to avoid the influence of outliers during normalization, SPECT V images were normalized to the median value, which is a relatively robust index, instead of the mean value of the whole-lung volumes. These rescaled or normalized data

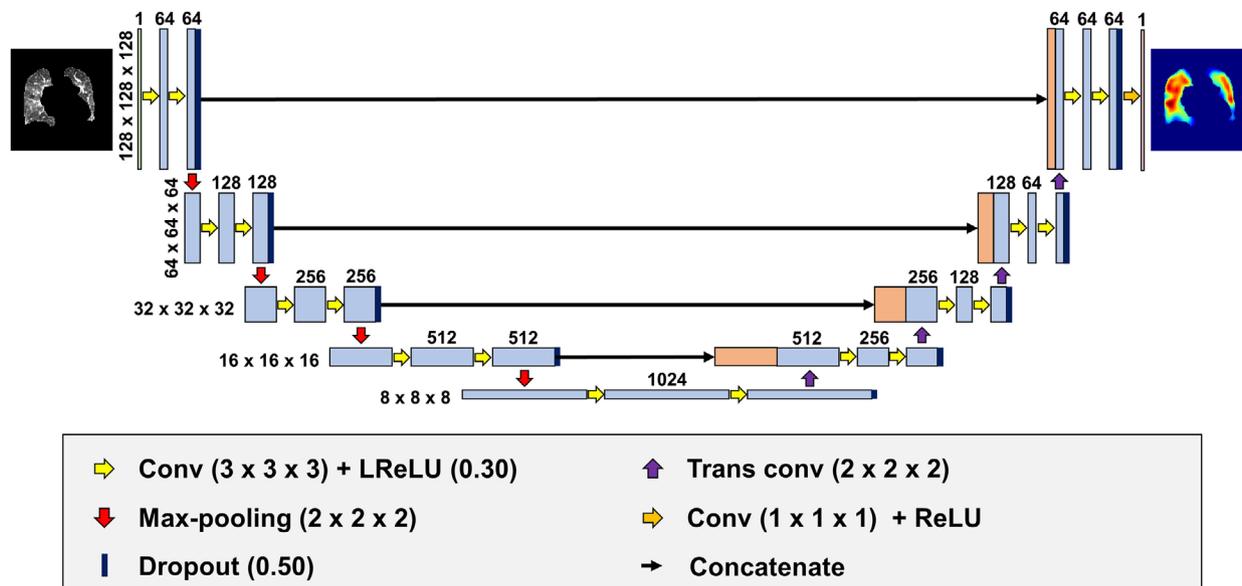


FIGURE 1 Model architecture used within the study. This CNN model is based on the U-Net architecture. The masked CT volume is input into the model, and the masked ventilation volume is output through the encoding and decoding paths. Numbers on the left side and top of the model represent the volume shape and the number of feature maps in a particular layer, respectively. CNN, convolutional neural network; Conv, convolutional layer; CT, computed tomography; Trans conv: transposed convolutional layer

volumes were replaced with 0 except for the whole-lung volumes shown as the input and output in Figure 1. The previously mentioned processes were implemented using Python 3.8.8.

2.3 | DL framework and training

We implemented a CNN model based on the U-Net architecture.²⁸ Figure 1 shows the model architecture. The architecture consists of encoding and decoding paths as with the general U-Net architecture; the encoding path abstracts and extracts low–high-dimensional features for input data, and the decoding path transforms the features to other domain output data. The encoding path consists of the repeated module of two $3 \times 3 \times 3$ convolutions with $1 \times 1 \times 1$ stride and $1 \times 1 \times 1$ padding, each followed by a leaky rectified linear unit (Leaky ReLU)²⁹ with $\alpha = 0.30$ and a connecting dropout layer³⁰ with a dropout rate = 0.50 and a $2 \times 2 \times 2$ max-pooling layer with a $2 \times 2 \times 2$ stride for down-sampling. At each down-sampling step, we doubled the number of feature maps. Every step in the decoding path consisted of a $2 \times 2 \times 2$ transposed convolutions with a $2 \times 2 \times 2$ stride that halved the number of feature maps, a concatenation with the feature maps with corresponding size from the encoding path, and two $3 \times 3 \times 3$ convolutions with a $1 \times 1 \times 1$ stride and $1 \times 1 \times 1$ padding, each followed by Leaky ReLU with $\alpha = 0.30$ and a connecting dropout layer with a dropout rate = 0.50. At the final layer, a $1 \times 1 \times 1$ convolution with a $1 \times 1 \times 1$ stride followed by a ReLU³¹ was used for output. L2 regularization was set to 1.0×10^{-4}

in all trainable layers. The 3D U-Net was implemented in Keras with a TensorFlow 2.3.0 backend.

In addition to the standard estimation method, an MCD approach was used to quantify prediction uncertainty. In contrast to the standard method without dropout during estimation, the MCD approach approximates the Gaussian process by repeatedly sampling with dropout during the estimation phase of the model's weights from a Bernoulli distribution²⁶ and could obtain a prediction uncertainty in each case.

To evaluate the model performance robustly, a total of 60 cases for training and validation were divided using a fivefold cross-validation (CV) procedure: 48 cases were used for training and 12 cases were used for validation in each fold. To keep the study as fair as possible, the model architecture and datasets were fixed in both approaches (with or without MCD) during the training process. Adam³² ($\text{lr} = 5.0 \times 10^{-5}$, $\beta_1 = 0.90$, $\beta_2 = 0.999$, $\varepsilon = 1.0 \times 10^{-7}$) was used for the optimizer, the maximum train epoch was set to 1000, the batch size was set to 2, and early stopping (patience = 50) was used. We used the mean squared error (MSE) as the loss function. Computation was performed using an Intel Core i9-9820X @ 3.30 Gz and an NVIDIA TITAN RTX GPU graphics card with a 24-GB memory.

2.4 | Evaluation

From these five trained models in each fold for both models with or without MCD, we then chose the best performance model on the basis of its validation loss (i.e., minimal error) and evaluated the model on the test set.

To test these final model performances for the model without MCD (standard model), we applied the test set to each trained model and averaged the test prediction results from the five trained models to acquire the mean test result (bagging) for each approach. For the model with MCD (MCD model), the model was used to predict repeatedly (sample of size = 200), and the average and SDs calculated in each voxel from the predicted results were the prediction results for each fold.

For quantitative analysis, we calculated Spearman's rank correlation coefficient (Spearman r_s) to assess the intensity correlation and the DSC to quantify the similarity of functional geometries for three separated nonoverlapping functional regions (i.e., high, moderate, and low) and MSE within the whole lung to evaluate the coincidence of raw intensity values. These functional regions were defined using the following thresholds: the highest 33.3 percentile value, lowest 33.3 percentile value, and between these values within the whole-lung volume. These functional regions' volumes were separated almost equally.

Spearman's r_s was defined by the following equation:

$$r_s = \frac{\sum_{i=1}^N [(y_i - \bar{y}) \cdot (p_i - \bar{p})]}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^N (p_i - \bar{p})^2}}, \quad (1)$$

where \bar{p} , \bar{y} , p_i , and y_i denote the average value and value at voxel i for the CTVI and SPECT V (ground truth), respectively. N denotes the total number of voxels in the whole lung.

DSC was defined by the following equation:

$$DSC(a, b) = \frac{2|a \cap b|}{|a| + |b|}, \quad (2)$$

where a denotes the volumes of functional lung regions in CTVI, and b denotes the volumes of the corresponding regions in SPECT V. Statistical differences were evaluated by the Wilcoxon signed-rank test significant difference procedure using SciPy v1.6.3.

To compare the existing approach, we calculated DSC between SPECT V low-functional regions and low-attenuation (emphysematous lesion) areas derived from 3DCT. These low-attenuation areas in 3DCT were defined as low-functional regions using the single upper threshold limit of -856 HU.²⁷ To keep the evaluation as fair as possible, SPECT V and CTVI images were segmented so that these volumes were almost equal to the low-attenuation volumes in 3DCT in each case.

Furthermore, to evaluate the influence of misregistration between CT and SPECT V, we calculated the DSC between two masks generated from CT and SPECT V. The CT masks have been described previously. The SPECT V masks were generated from SPECT V images using a lower threshold limit of 15% of the maximum intensity value of the images,³³ and they were modi-

fied manually if necessary (e.g., deleting obvious airway deposition). In addition, we evaluated Spearman r_s between the DSC calculated for these two masks and the quantification indices between SPECT V and CTVI.

For evaluating prediction uncertainty, SDs were calculated. Furthermore, to minimize the influence of large or small prediction values, the voxel-wise coefficient of variation map was defined as the scaled uncertainty map, and the SDs were divided by the average values per voxel. For quantitative evaluation, the average values in three functional regions were calculated.

3 | RESULTS

3.1 | Fivefold cross-validation

Table 2 summarizes the results of the voxel-wise Spearman r_s and DSCs in the high-, moderate-, and low-functional regions (DSC_{high} , $DSC_{moderate}$, DSC_{low} , respectively) obtained between CTVI with MCD (CTVI_{MCD U-Net}) and SPECT V and between CTVI without MCD (CTVI_{U-Net}) and SPECT V within the whole-lung volume for 11 test cases. For the average value of each fold, the average similarity metrics with 1SD between CTVI_{MCD U-Net} and SPECT V were 0.73 ± 0.07 , 0.67 ± 0.07 , 0.48 ± 0.06 , and 0.73 ± 0.05 for the Spearman r_s , DSC_{high} , $DSC_{moderate}$, and DSC_{low} , respectively. The average similarity metrics with 1SD between CTVI_{U-Net} and SPECT V were 0.71 ± 0.06 , 0.65 ± 0.04 , 0.47 ± 0.04 , and 0.73 ± 0.06 for the Spearman r_s , DSC_{high} , $DSC_{moderate}$, and DSC_{low} , respectively. Compared with CTVI_{U-Net}, CTVI_{MCD U-Net} exhibited better performance in every evaluation index and showed comparable or superior performance in each fold with no statistically significant difference, excluding Spearman r_s and DSC_{high} in the second fold. Moreover, for the average value of each fold, the average MSE with 1SD values were 0.11 ± 0.04 and 0.15 ± 0.05 for CTVI_{MCD U-Net} and CTVI_{U-Net}, respectively. Compared with CTVI_{U-Net}, CTVI_{MCD U-Net} exhibited better performance in every evaluation index and showed superior performance in each fold with statistically significant difference, excluding those in the fifth fold.

3.2 | Bagging performance evaluation

Figures 2 and 3 show SPECT V, CTVI_{MCD U-Net}, CTVI_{U-Net}, masked CT coronal-plane images, and scatter plots between SPECT V and each CTVI within whole-lung volume for typical patients. The figures show two examples: one with relatively good agreement between DL-based CTVI and SPECT V and the other with relatively poor agreement between DL-based CTVI and SPECT V. Compared with CTVI_{U-Net}, CTVI_{MCD U-Net} showed greater similarity with SPECT V. The low-functional region (orange arrows) was predicted better

TABLE 2 The prediction performance for the 11 test cases for each of the 5 models in each model with or without MCD (MCD U-Net and U-Net) trained by fivefold cross-validation

		Spearman r_s	Dice similarity coefficient		
			High	Moderate	Low
First fold	CTVI _{MCD U-Net}	0.73 ± 0.07	0.67 ± 0.08	0.48 ± 0.06	0.73 ± 0.05
	CTVI _{U-Net}	0.70 ± 0.06	0.64 ± 0.04	0.46 ± 0.04	0.73 ± 0.06
Second fold	CTVI _{MCD U-Net}	0.74 ± 0.05	0.68 ± 0.06	0.49 ± 0.05	0.74 ± 0.04
	CTVI _{U-Net}	0.70 ± 0.07	0.63 ± 0.04	0.45 ± 0.04	0.73 ± 0.06
Third fold	CTVI _{MCD U-Net}	0.74 ± 0.06	0.67 ± 0.07	0.48 ± 0.06	0.74 ± 0.05
	CTVI _{U-Net}	0.73 ± 0.05	0.65 ± 0.05	0.47 ± 0.04	0.75 ± 0.04
Fourth fold	CTVI _{MCD U-Net}	0.72 ± 0.07	0.67 ± 0.08	0.47 ± 0.06	0.72 ± 0.05
	CTVI _{U-Net}	0.71 ± 0.06	0.64 ± 0.04	0.46 ± 0.04	0.73 ± 0.06
Fifth fold	CTVI _{MCD U-Net}	0.73 ± 0.07	0.68 ± 0.06	0.48 ± 0.04	0.73 ± 0.05
	CTVI _{U-Net}	0.72 ± 0.06	0.66 ± 0.04	0.47 ± 0.04	0.73 ± 0.06
Average	CTVI _{MCD U-Net}	0.73 ± 0.07	0.67 ± 0.07	0.48 ± 0.06	0.73 ± 0.05
	CTVI _{U-Net}	0.71 ± 0.06	0.65 ± 0.04	0.47 ± 0.04	0.73 ± 0.06

Abbreviations: CTVI, computed tomography ventilation imaging; MCD, Monte Carlo dropout.

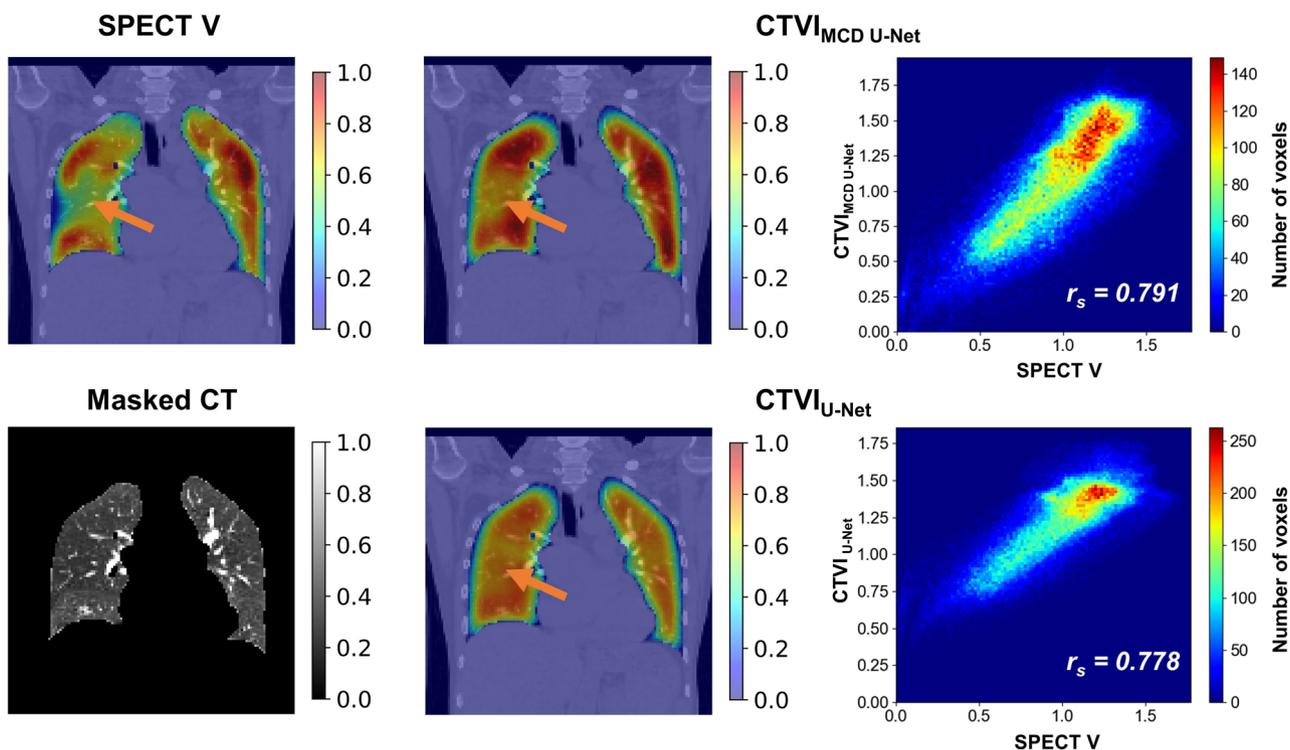


FIGURE 2 Typical cases of coronal slices showing relatively good performance based on the Spearman r_s values between SPECT V and CTVIs of both models. The top row shows the SPECT V on the left, followed by the ventilation predictions of the CTVI_{MCD U-Net} and the scatter plot between SPECT V and CTVI_{MCD U-Net} within whole-lung volume. The bottom row shows the masked CT on the left, followed by the ventilation predictions of the CTVI_{U-Net} and the scatter plot between SPECT V and CTVI_{U-Net} within whole-lung volume. Orange arrows indicate the defect regions of SPECT V. The ventilation values for viewing were normalized to give the 99th percentile value of 1 and the 1st percentile value of 0. CT, computed tomography; CTVI, computed tomography ventilation imaging; MCD, Monte Carlo dropout; SPECT V, single-photon emission computed tomography ventilation

by CTVI_{MCD U-Net} than by CTVI_{U-Net}. Although most of the whole lung showed good agreement, both models with or without MCD tended toward functional overestimation.

Figure 4 compares the Spearman rank correlation coefficients and DSCs for three functional regions between CTVI_{MCD U-Net} and CTVI_{U-Net} methods and SPECT V for the 11 test cases. The average similarity

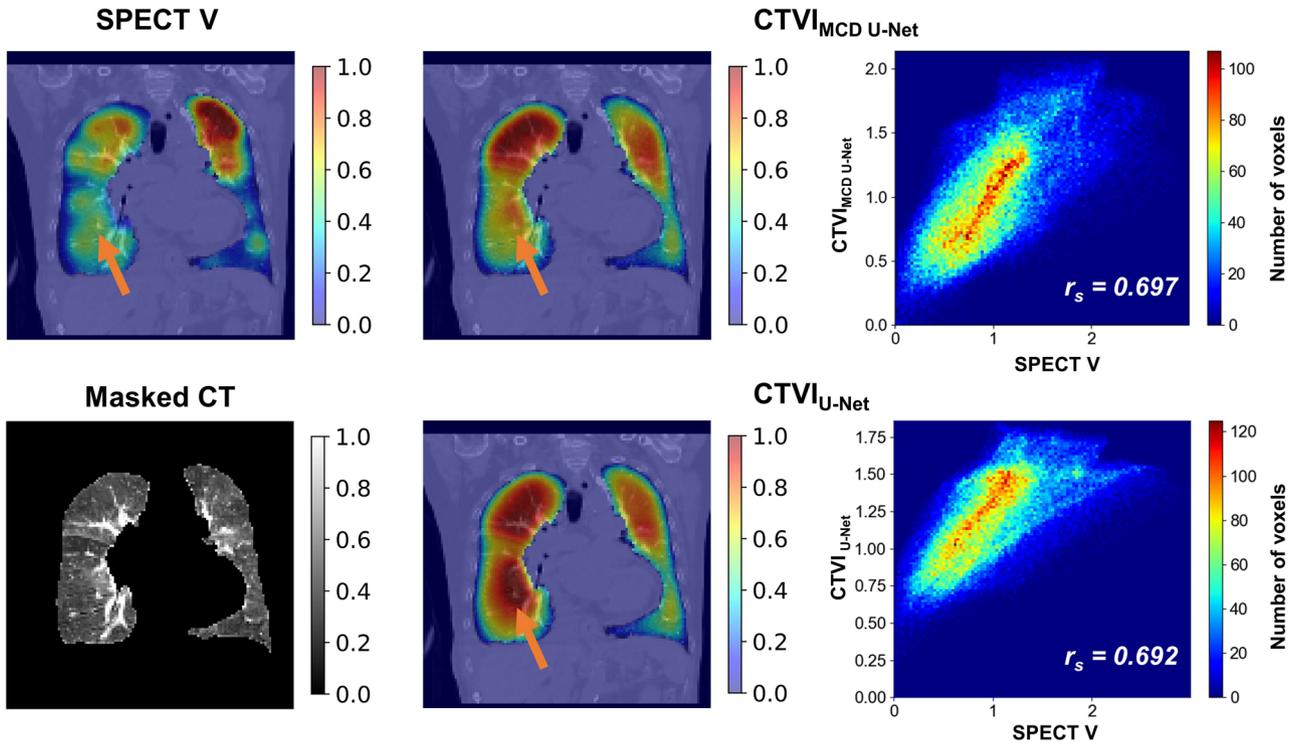


FIGURE 3 Typical cases of coronal slices showing relatively poor performances based on the Spearman r_s values between SPECT V and CTIVs of both models. The top row shows the SPECT V on the left side, followed by the ventilation predictions of the $CTVI_{MCD U-Net}$ and the scatter plot between SPECT V and $CTVI_{MCD U-Net}$ within whole-lung volume. The bottom row shows the masked CT on the left side, followed by the ventilation predictions of the $CTVI_{U-Net}$ and the scatter plot between SPECT V and $CTVI_{U-Net}$ within whole-lung volume. Orange arrows indicate the defect regions of SPECT V. The ventilation values for viewing were normalized to give the 99th percentile value of 1 and the 1st percentile value of 0. CT, computed tomography; CTVI, computed tomography ventilation imaging, MCD, Monte Carlo dropout; SPECT V, single-photon emission computed tomography ventilation

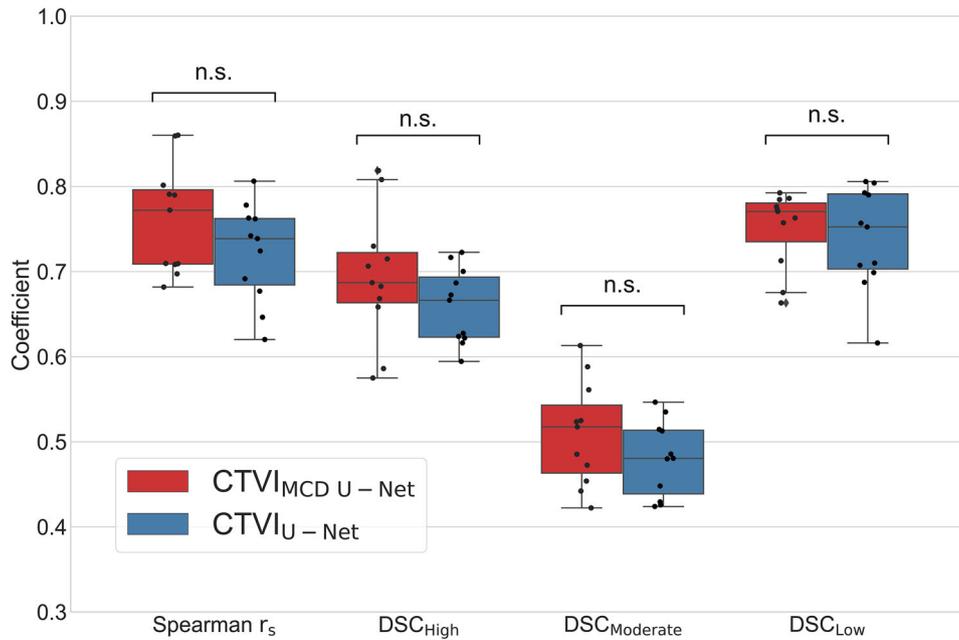


FIGURE 4 Boxplot and strip plot showing the bagged prediction performance for the 11 test cases for each of the 5 CNN models (MCD U-Net and U-Net) trained by fivefold cross-validation. The Spearman r_s values are shown on the left, followed by the DSC_{high} , $DSC_{moderate}$, and DSC_{low} values between SPECT V and the two DL models. * $p < 0.001$. CNN, convolutional neural network; DSC, dice similarity coefficient; MCD, Monte Carlo dropout; n.s., not significant; SPECT V, single-photon emission computed tomography ventilation

TABLE 3 Prediction performance for the bagging and fivefold CV averaging results for 11 test cases for each of the 5 CNN models (MCD U-Net and U-Net) trained by fivefold CV

	Method	Spearman r_s	Dice similarity coefficient		
			High	Moderate	Low
CTVI _{MCD U-Net}	Bagging	0.76 ± 0.06	0.69 ± 0.07	0.51 ± 0.06	0.75 ± 0.04
	Average CV	0.73 ± 0.07	0.67 ± 0.07	0.48 ± 0.06	0.73 ± 0.05
CTVI _{U-Net}	Bagging	0.72 ± 0.05	0.66 ± 0.04	0.48 ± 0.04	0.74 ± 0.06
	Average CV	0.71 ± 0.06	0.65 ± 0.04	0.47 ± 0.04	0.73 ± 0.06

Abbreviations: CNN, convolutional neural network; CTVI, computed tomography ventilation imaging; CV, cross-validation; MCD, Monte Carlo dropout.

metrics with 1SD between CTVI_{MCD U-Net} and SPECT V were 0.76 ± 0.06 , 0.69 ± 0.07 , 0.51 ± 0.06 , and 0.75 ± 0.04 for the Spearman r_s , DSC_{high} , $DSC_{moderate}$, and DSC_{low} , respectively. The average similarity metrics with 1SD between CTVI_{U-Net} and SPECT V were 0.72 ± 0.05 , 0.66 ± 0.04 , 0.48 ± 0.04 , and 0.74 ± 0.06 for the Spearman r_s , DSC_{high} , $DSC_{moderate}$, and DSC_{low} , respectively. Compared with CTVI_{U-Net}, CTVI_{MCD U-Net} had better performance with no statistically significant difference. Furthermore, the average MSE with 1SD values were 0.09 ± 0.04 and 0.12 ± 0.05 for CTVI_{MCD U-Net} and CTVI_{U-Net}, respectively. Compared with CTVI_{U-Net}, CTVI_{MCD U-Net} exhibited better performance with statistically significant difference. The DSC values for SPECT V were 0.48 ± 0.19 , 0.47 ± 0.17 , and 0.10 ± 0.07 for CTVI_{MCD U-Net}, CTVI_{U-Net}, and low-attenuation areas in 3DCT, respectively. Therefore, compared with low-attenuation areas, both CTVI_{MCD U-Net} and CTVI_{U-Net} showed better performance. Moreover, the Spearman r_s between DSC for the two masks (i.e., generated from CT or SPECT V) and the quantification indices between CTVI_{MCD U-Net} and SPECT V were 0.19, 0.23, 0.35, and 0.02 for Spearman r_s , DSC_{high} , $DSC_{moderate}$, and DSC_{low} , respectively. Conversely, the Spearman r_s between DSC from these two masks and the quantification index between CTVI_{U-Net} and SPECT V were 0.14, 0.32, 0.30, and -0.15 for Spearman r_s , DSC_{high} , $DSC_{moderate}$, and DSC_{low} , respectively. There was no tendency toward significantly worse prediction accuracy for the cases with relatively poor registration between CT and SPECT.

Table 3 summarizes the results of the voxel-wise Spearman r_s values and DSCs in the high-, moderate-, and low-functional regions obtained between the CTVI_{MCD U-Net} and CTVI_{U-Net} methods and SPECT V within the whole-lung volume for 11 test cases for comparison with bagging and CV average results. The best performance of each result was CTVI_{MCD U-Net} with bagging, followed by CTVI_{MCD U-Net} with averaged CV results, CTVI_{U-Net} with bagging, and CTVI_{U-Net} with averaged CV results in descending order. The range of improvement of the bagging technique between the averaged CV results was larger for CTVI_{MCD U-Net} than for CTVI_{U-Net}.

3.3 | Uncertainty evaluation

Figure 5 shows the SPECT V, CTVI_{MCD U-Net}, scaled uncertainty, and masked CT images for a typical case in the coronal plane. The figure shows a tendency toward larger uncertainty on the lung edge (e.g., diaphragm) than in the center and larger uncertainty in the dorsal than ventral lung. The average SD values within the whole-lung volume were 0.40 ± 0.01 , 0.30 ± 0.01 , and 0.19 ± 0.01 in the high-, moderate-, and low-functional regions, respectively. Conversely, the average scaled uncertainties within the whole-lung volume were 0.27 ± 0.00 , 0.27 ± 0.01 , and 0.36 ± 0.03 in the high-, moderate-, and low-functional regions, respectively. Compared with the high- and moderate-functional regions, the low-functional region showed larger scaled uncertainty.

4 | DISCUSSION

In this study, we evaluated the DL-based framework that translates 3DCT images to lung-ventilation images directly. Our DL-based model had moderate or high performance, indicating that the DL-based approach was effective for lung-ventilation estimation mapping from 3DCT. To the best of our knowledge, this is the first study to evaluate the DL-based approach for transforming 3DCT to SPECT V.

Our results showed that the DL approach had moderate or high performance for lung-ventilation estimation; the average similarity metrics with 1SD between CTVI_{MCD U-Net} and SPECT V were 0.76 ± 0.06 , 0.69 ± 0.07 , 0.51 ± 0.06 , and 0.75 ± 0.04 for Spearman r_s , DSC_{high} , $DSC_{moderate}$, and DSC_{low} , respectively. Liu et al. implemented the DL-based approach to transform 4DCT (two or ten phases) to ^{99m}Tc-gas SPECT V images.²³ They showed that use of 10 phases had better performance than 2 phases. They showed that the average similarity metrics with 1SD when using 10/2 phases were $0.73 \pm 0.16/0.71 \pm 0.17$, $0.83 \pm 0.07/0.83 \pm 0.07$, $0.61 \pm 0.10/0.59 \pm 0.10$, and $0.74 \pm 0.10/0.72 \pm 0.10$ for Spearman r_s , DSC_{high} , $DSC_{moderate}$, and DSC_{low} , respectively. Although the trend of our results was similar to that of their results, our results were

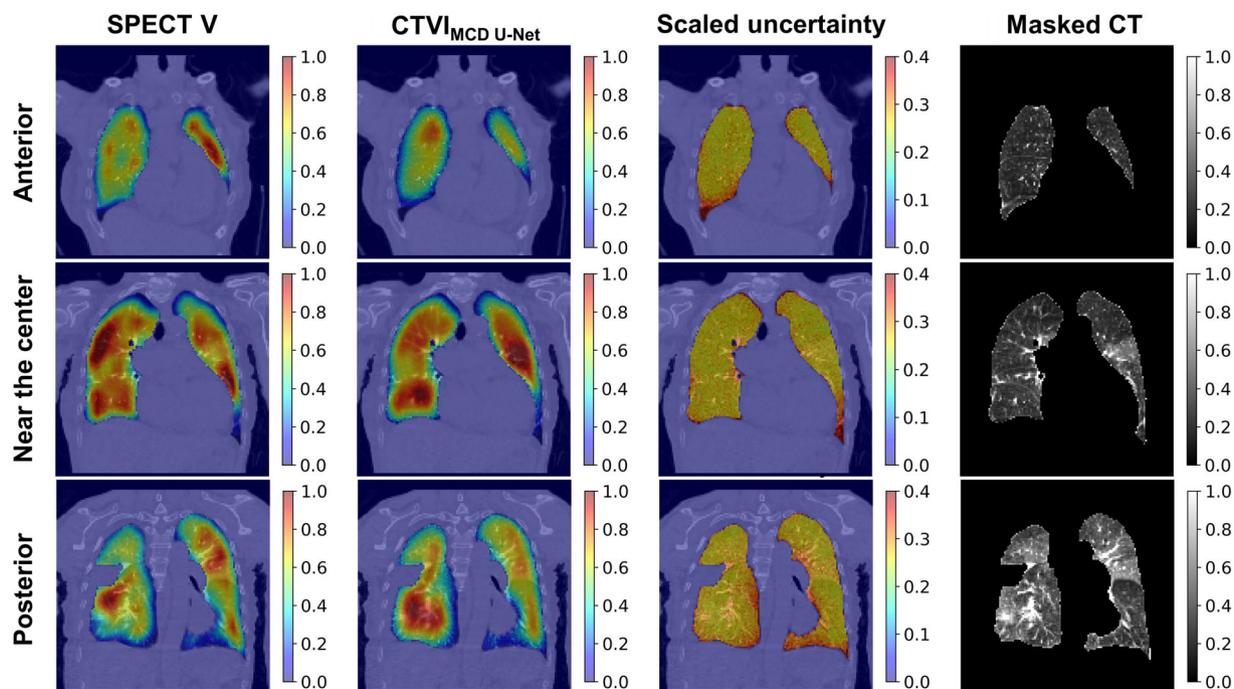


FIGURE 5 Typical cases of prediction of $CTVI_{MCD\ U-Net}$ in coronal slices. The anterior slice is shown on the top row, followed by the near, center, and posterior slices. The SPECT V is shown on the left, followed by $CTVI_{MCD\ U-Net}$, the scaled uncertainty for the $CTVI_{MCD\ U-Net}$, and masked CT. The ventilation values for viewing were normalized to give the 99th percentile value of 1 and 1st percentile value of 0. CT, computed tomography; CTVI, computed tomography ventilation imaging; MCD, Monte Carlo dropout; SPECT V, single-photon emission computed tomography ventilation

slightly inferior to theirs, which could have been caused by the following two factors: First, it is possible that 4DCT images contain more physiological and biological information necessary for estimating lung-ventilation function than 3DCT images. Second, the differences in patient characteristics (patients with mainly circulatory disease vs. patients with esophagus or lung cancer) could have affected the results. Ren et al. implemented the DL-based approach to transform 3DCT to SPECT perfusion images and investigated the influence of different model architectures and preprocessing.^{24,25} They reported that there were some cases in which the low-functional region was predicted to be a high-functional region, which can be partly attributed to the low occurrence and small volume of the low-functional regions compared with high-functional regions. Our results showed similar trends (Figures 3 and 4). The percentage of non-defect ventilation volume segmented and calculated in our SPECT V images was $68.0\% \pm 11.1\%$, and our datasets had larger non-defect ventilation regions than defect regions. Thus, our DL model might have been affected by the distributions of ventilation regions and tended to estimate low-functional region to be a high-functional region.

This study showed that $CTVI_{MCD\ U-Net}$ had better predictive performance with the bagging technique than $CTVI_{U-Net}$: 0.76 ± 0.06 versus 0.72 ± 0.05 for Spearman r_s , 0.69 ± 0.07 versus 0.66 ± 0.04 for DSC_{high} , $0.51 \pm$

0.06 versus 0.48 ± 0.04 for $DSC_{moderate}$, and 0.75 ± 0.04 versus 0.74 ± 0.06 for DSC_{low} . The $DSC_{moderate}$ is always worse compared with DSC_{high} and DSC_{low} . It could be caused by a technical issue in the evaluation method. We selected the evaluation approach with three nonoverlapping functional regions (i.e., high, moderate, and low) divided equally in volume based on threshold limits. This approach has generally been used in lung-ventilation studies,^{34,35} is suggested for use in functional lung-avoidance radiotherapy treatment planning,³⁶ and shows a similar tendency. In this approach, high- and low-functional regions were defined by a single lower/upper threshold limit. Conversely, moderate-functional regions were defined by two threshold limits. Therefore, moderate-functional regions might have been more influenced by the border uncertainty compared with high- and low-functional regions; therefore, the DSC value in moderate-functional regions might be always the worst.

In addition, $CTVI_{MCD\ U-Net}$ showed an increased range of performance between bagging and fivefold CV averaged results compared to $CTVI_{U-Net}$: 0.03 versus 0.01 for Spearman r_s , 0.02 versus 0.01 for DSC_{high} , 0.03 versus 0.01 for $DSC_{moderate}$, and 0.02 versus 0.01 for DSC_{low} . The difference in performance could be caused by the difference in influences of the ensemble effect. The number of samples for estimation with the MCD model was set to 200 in each fold, bagging (i.e.,

averaging) the prediction results of 5 models from CV. Therefore, the influence of the ensemble effect was greater for the MCD model than for the standard model (1000 models vs. 5 models). Furthermore, these influences were observed for fivefold CV averaging results.

The scaled uncertainties were 0.27 ± 0.00 , 0.27 ± 0.01 , and 0.36 ± 0.03 in the high-, moderate-, and low-functional regions, respectively. Compared with the high- and moderate-functional regions, the low-functional regions had larger scaled uncertainties. As mentioned previously, the low-functional regions could not be estimated well because many patients showed normal function, and the number of patients showing deficits was small. In addition, as shown in Figure 5, the uncertainty tended to be large at the lung periphery and near the diaphragm. The blurring around the diaphragm was particularly large and could have been caused by breathing and heartbeat as well as by the variability in lung shapes among the patients. The blurring problem might be mitigated by aligning the sizes of pulmonary parenchyma in SPECT and CT images well. Furthermore, the improvement of SPECT images using respiratory-gated motion correction³⁷ during image acquisition might be effective.

Although our results showed the potential of DL-based approach for lung-ventilation estimation mapping, there were several limitations in this study. First, our datasets have bias in patient background. Generally, DL is strongly affected by the bias of the datasets. Although our results showed that the DL-based model had moderate or high performance in this study, it is unknown whether the models trained by our datasets can be adapted for those who are undergoing radiotherapy and different patient backgrounds. Our datasets were different from clinical population receiving radiotherapy generally (e.g., in terms of non-defect ventilation volume and gender). The percentage of the non-defect ventilation volume was $68.0\% \pm 11.1\%$, and the datasets had a larger percentage of non-defect ventilation volume than the clinical population with early-stage primary non-small-cell lung cancer ($53.7\% \pm 15.5\%$) in the previous study.¹² This occurred because our datasets consisted of cases showing non-defect ventilation regions (e.g., pulmonary hypertension and pulmonary embolism). In addition, the datasets had gender imbalance because of the bias of the clinical population in our hospital patient cohort, as it is different from the clinical population (male > female) receiving radiotherapy treatments such as for lung cancer. Further studies are necessary to investigate the effect of data characteristics, and we will expand the data pool with more datasets. Second, the model architecture and hyper parameters were not completely optimized for the lung-ventilation estimation mapping task. In this study, the model architecture was based on the standard U-Net model for image generation in medical imaging, and the hyper param-

eters were set to general parameters (i.e., optimizer: Adam, dropout rate: 0.50). Although we only used a relatively general model and hyper parameters, we did not consider any optimization, such as adding layers or tuning hyper parameters. Therefore, the optimization of the model architecture and hyper parameters may improve the estimation performance. Third, positional errors between SPECT V and 3DCT were observed in our dataset. Although the 3DCT and SPECT V images were acquired during the same examination series, differences in patient positions and/or anatomical geometries (e.g., lung shape and diaphragm position) between SPECT V and 3DCT were observed because of changes in the breathing patterns during SPECT and/or 3DCT imaging. To minimize the influence of these errors, we registered SPECT V and 3DCT images rigidly at preprocessing, and the registration accuracy for each patient was checked visually. However, the influences of changes in breathing patterns between the SPECT V and 3DCT images were not completely eliminated, and there was some possibility that data were affected in the training and/or prediction process by these mismatches. Further studies are necessary to investigate the influence of these mismatches. Finally, the approach for uncertainty quantification employed in this study was the simplest method (i.e., calculating only the SD). Therefore, we did not calibrate the quality of uncertainty quantification and only calculated SDs as epistemic uncertainty (not aleatoric uncertainty). Although this approach has been used in previous medical-imaging studies,^{38,39} it has some limitations as described previously. Further studies are necessary to estimate more sophisticated uncertainty quantification for DL-based frameworks that translate lung-ventilation images from 3DCT images.

This pulmonary ventilation mapping prediction tool might be useful for adaptive lung-functional avoidance radiotherapy planning. Yuan et al. reported that changes in pulmonary regional ventilation and perfusion functions were sometimes observed by SPECT V/Q during the course of radiotherapy because of tumor shrinkage.⁴⁰ Kipritidis et al. also reported inter-fractional changes in pulmonary regional ventilation functions captured by CTVI based on cone-beam (CB) CT over a course of radiotherapy, supporting the feasibility for adaptive lung-ventilation-avoidance radiotherapy.⁴¹ Our results showed that the DL-based approach was effective for lung-ventilation estimation mapping from CT. Therefore, this approach could be potentially applied for CBCT to lung-ventilation translation. CBCT is often acquired for positional alignment during the course of radiation treatment, and it might be possible to monitor these inter-fractional pulmonary functional changes in a simpler way. This DL-based approach could also be potentially used for adaptive functional lung-avoidance radiotherapy and treatment response modeling.

5 | CONCLUSIONS

We evaluated the performance and uncertainty of DL-based frameworks that translated 3D lung-ventilation images from 3DCT images. Our results suggested that the DL-based approach was effective for lung-ventilation estimation mapping from 3DCT.

ACKNOWLEDGMENT

This study was supported in part by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (C) (20K16733).

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

1. Yaremko BP, Guerrero TM, Noyola-Martinez J, et al. Reduction of normal lung irradiation in locally advanced non-small-cell lung cancer patients, using ventilation images for functional avoidance. *Int J Radiat Oncol Biol Phys*. 2007;68(2):562–571. <https://doi.org/10.1016/j.ijrobp.2007.01.044>
2. Faught AM, Miyasaka Y, Kadoya N, et al. Evaluating the toxicity reduction with computed tomographic ventilation functional avoidance radiation therapy. *Int J Radiat Oncol Biol Phys*. 2017;99(2):325–333. <https://doi.org/10.1016/j.ijrobp.2017.04.024>
3. Sawant A, Yamamoto T, Cai J. Treatment planning based on lung functional avoidance is not ready for clinical deployment. *Med Phys*. 2018;45(6):2353–2356. <https://doi.org/10.1002/mp.12881>
4. Vinogradskiy Y, Castillo R, Castillo E, et al. Use of 4-dimensional computed tomography-based ventilation imaging to correlate lung dose and function with clinical outcomes. *Int J Radiat Oncol Biol Phys*. 2013;86(2):366–371. <https://doi.org/10.1016/j.ijrobp.2013.01.004>
5. Yamamoto T, Kabus S, Bal M, Keall P, Benedict S, Daly M. The first patient treatment of computed tomography ventilation functional image-guided radiotherapy for lung cancer. *Radiother Oncol*. 2016;118(2):227–231. <https://doi.org/10.1016/j.radonc.2015.11.006>
6. Hoover DA, Capaldi DP, Sheikh K, et al. Functional lung avoidance for individualized radiotherapy (FLAIR): study protocol for a randomized, double-blind clinical trial. *BMC Cancer*. 2014;14:934. <https://doi.org/10.1186/1471-2407-14-934>
7. Petersson J, Sanchez-Crespo A, Larsson SA, Mure M. Physiological imaging of the lung: single-photon-emission computed tomography (SPECT). *J Appl Physiol (1985)*. 2007;102(1):468–476. [10.1152/jappphysiol.00732.2006](https://doi.org/10.1152/jappphysiol.00732.2006)
8. Kipritidis J, Siva S, Hofman MS, Callahan J, Hicks RJ, Keall PJ. Validating and improving CT ventilation imaging by correlating with ventilation 4D-PET/CT using 68Ga-labeled nanoparticles. *Med Phys*. 2014;41(1):011910. <https://doi.org/10.1118/1.4856055>
9. van Beek EJ, Wild JM, Kauczor HU, Schreiber W, Mugler JP 3rd, de Lange EE. Functional MRI of the lung using hyperpolarized 3-helium gas. *J Magn Reson Imaging*. 2004;20(4):540–554. <https://doi.org/10.1002/jmri.20154>
10. Chae EJ, Seo JB, Goo HW, et al. Xenon ventilation CT with a dual-energy technique of dual-source CT: initial experience. *Radiology*. 2008;248(2):615–624. <https://doi.org/10.1148/radiol.2482071482>
11. Guerrero T, Sanders K, Noyola-Martinez J, et al. Quantification of regional ventilation from treatment planning CT. *Int J Radiat Oncol Biol Phys*. 2005;62(3):630–634. <https://doi.org/10.1016/j.ijrobp.2005.03.023>
12. Hegi-Johnson F, Keall P, Barber J, Bui C, Kipritidis J. Evaluating the accuracy of 4D-CT ventilation imaging: first comparison with Technegas SPECT ventilation. *Med Phys*. 2017;44(8):4045–4055. <https://doi.org/10.1002/mp.12317>
13. Vinogradskiy Y, Jackson M, Schubert L, et al. Assessing the use of 4DCT-ventilation in pre-operative surgical lung cancer evaluation. *Med Phys*. 2017;44(1):200–208. <https://doi.org/10.1002/mp.12026>
14. Kipritidis J, Hofman MS, Siva S, et al. Estimating lung ventilation directly from 4D CT Hounsfield unit values. *Med Phys*. 2016;43(1):33. <https://doi.org/10.1118/1.4937599>
15. Sahiner B, Pezeshk A, Hadjiiski LM, et al. Deep learning in medical imaging and radiation therapy. *Med Phys*. 2019;46(1):e1–e36. <https://doi.org/10.1002/mp.13264>
16. Zhu W, Huang Y, Zeng L, et al. AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med Phys*. 2019;46(2):576–589. <https://doi.org/10.1002/mp.13300>
17. Nguyen D, Jia X, Sher D, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. *Phys Med Biol*. 2019;64(6):065020. <https://doi.org/10.1088/1361-6560/ab039b>
18. Kajikawa T, Kadoya N, Ito K, et al. A convolutional neural network approach for IMRT dose distribution prediction in prostate cancer patients. *J Radiat Res*. 2019;60(5):685–693. <https://doi.org/10.1093/jrr/rrz051>
19. Kontaxis C, Bol GH, Lagendijk JJW, Raaymakers BW. DeepDose: towards a fast dose calculation engine for radiation therapy using deep learning. *Phys Med Biol*. 2020;65(7):075013. <https://doi.org/10.1088/1361-6560/ab7630>
20. Kimura Y, Kadoya N, Tomori S, Oku Y, Jingu K. Error detection using a convolutional neural network with dose difference maps in patient-specific quality assurance for volumetric modulated arc therapy. *Phys Med*. 2020;73:57–64. <https://doi.org/10.1016/j.ejmp.2020.03.022>
21. Kimura Y, Kadoya N, Oku Y, Kajikawa T, Tomori S, Jingu K. Error detection model developed using a multi-task convolutional neural network in patient-specific quality assurance for volumetric-modulated arc therapy. *Med Phys*. 2021;48(9):4769–4783. <https://doi.org/10.1002/mp.15031>
22. Zhong Y, Vinogradskiy Y, Chen L, et al. Technical Note: deriving ventilation imaging from 4DCT by deep convolutional neural network. *Med Phys*. 2019;46(5):2323–2329. <https://doi.org/10.1002/mp.13421>
23. Liu Z, Miao J, Huang P, et al. A deep learning method for producing ventilation images from 4DCT: first comparison with Technegas SPECT ventilation. *Med Phys*. 2020;47(3):1249–1257. <https://doi.org/10.1002/mp.14004>
24. Ren G, Zhang J, Li T, et al. Deep learning-based computed tomography perfusion mapping (DL-CTPM) for pulmonary CT-to-perfusion translation. *Int J Radiat Oncol Biol Phys*. 2021;110(5):1508–1518. <https://doi.org/10.1016/j.ijrobp.2021.02.032>
25. Ren G, Lam SK, Zhang J, et al. Investigation of a novel deep learning-based computed tomography perfusion mapping framework for functional lung avoidance radiotherapy. *Front Oncol*. 2021;11:644703. [10.3389/fonc.2021.644703](https://doi.org/10.3389/fonc.2021.644703)
26. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. 2016;11050–11059. arXiv:1506.02142.
27. Uchida Y, Tsugawa T, Tanaka-Mizuno S, et al. Exclusion of emphysematous lung from dose-volume estimates of risk improves prediction of radiation pneumonitis. *Radiat Oncol*. 2017;12(1):160. [10.1186/s13014-017-0891-z](https://doi.org/10.1186/s13014-017-0891-z)

28. Ronneberger O, Fischer P, Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer; 2015:234–241.
29. Maas AL, Hannun AY, Ng AY. *Rectifier Nonlinearities Improve Neural Network Acoustic Models*. CiteSeer; 2013:3.
30. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(1):1929–1958.
31. Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. Paper presented at: ICML. 2010.
32. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980.
33. Cheimariotis GA, Al-Mashat M, Haris K, et al. Automatic lung segmentation in functional SPECT images using active shape models trained on reference lung shapes from CT. *Ann Nucl Med*. 2018;32(2):94–104. <https://doi.org/10.1007/s12149-017-1223-y>
34. Yamamoto T, Kabus S, von Berg J, et al. Reproducibility of four-dimensional computed tomography-based lung ventilation imaging. *Acad Radiol*. 2012;19(12):1554–1565. <https://doi.org/10.1016/j.acra.2012.07.006>
35. Rankine LJ, Wang Z, Driehuys B, Marks LB, Kelsey CR, Das SK. Correlation of regional lung ventilation and gas transfer to red blood cells: implications for functional-avoidance radiation therapy planning. *Int J Radiat Oncol Biol Phys*. 2018;101(5):1113–1122. <https://doi.org/10.1016/j.ijrobp.2018.04.017>
36. Yamamoto T, Kabus S, von Berg J, Lorenz C, Keall PJ. Impact of four-dimensional computed tomography pulmonary ventilation imaging-based functional avoidance for lung cancer radiotherapy. *Int J Radiat Oncol Biol Phys*. 2011;79(1):279–288. <https://doi.org/10.1016/j.ijrobp.2010.02.008>
37. Ue H, Haneishi H, Iwanaga H, Suga K. Nonlinear motion correction of respiratory-gated lung SPECT images. *IEEE Trans Med Imaging*. 2006;25(4):486–495. <https://doi.org/10.1109/TMI.2006.871546>
38. Wickstrom K, Kampffmeyer M, Jenssen R. Uncertainty and interpretability in convolutional neural networks for semantic segmentation of colorectal polyps. *Med Image Anal*. 2020;60:101619. <https://doi.org/10.1016/j.media.2019.101619>
39. Nguyen D, Sadeghnejad Barkousaraie A, Bohara G, et al. A comparison of Monte Carlo dropout and bootstrap aggregation on the performance and uncertainty estimation in radiation therapy dose prediction with deep learning neural networks. *Phys Med Biol*. 2021;66(5):054002. <https://doi.org/10.1088/1361-6560/abe04f>
40. Yuan ST, Frey KA, Gross MD, et al. Changes in global function and regional ventilation and perfusion on SPECT during the course of radiotherapy in patients with non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys*. 2012;82(4):e631–e638. <https://doi.org/10.1016/j.ijrobp.2011.07.044>
41. Kipritidis J, Hugo G, Weiss E, Williamson J, Keall PJ. Measuring interfraction and intrafraction lung function changes during radiation therapy using four-dimensional cone beam CT ventilation imaging. *Med Phys*. 2015;42(3):1255–1267. <https://doi.org/10.1118/1.4907991>

How to cite this article: Kajikawa T, Kadoya N, Maehara Y, et al. A deep learning method for translating 3DCT to SPECT ventilation imaging: First comparison with $^{81\text{m}}\text{Kr}$ -gas SPECT ventilation imaging. *Med Phys*. 2022;49:4353–4364. <https://doi.org/10.1002/mp.15697>