



Lung Cancer Segmentation With Transfer Learning: Usefulness of a Pretrained Model Constructed From an Artificial Dataset Generated Using a Generative Adversarial Network

Mizuho Nishio^{1,2*}, Koji Fujimoto^{1,3}, Hidetoshi Matsuo², Chisako Muramatsu⁴, Ryo Sakamoto^{1,5} and Hiroshi Fujita⁶

¹Department of Diagnostic Imaging and Nuclear Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, ²Department of Radiology, Kobe University Hospital, Kobe, Japan, ³Department of Real World Data Research and Development, Kyoto University Graduate School of Medicine, Hikone, Japan, ⁴Faculty of Data Science, Shiga University, Gifu, Japan, ⁵Preemptive Medicine and Lifestyle-Related Disease Research Center, Kyoto University Hospital, Kyoto, Japan, ⁶Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering, Gifu University, Gifu, Japan

OPEN ACCESS

Edited by:

Maria F. Chan,
Memorial Sloan Kettering Cancer
Center, United States

Reviewed by:

Shivanand Sharanappa Gornale,
Rani Channamma University, India
Shailesh Tripathi,
Tampere University of Technology,
Finland

*Correspondence:

Mizuho Nishio
nishiomizuho@gmail.com

Specialty section:

This article was submitted to
Medicine and Public Health,
a section of the journal
Frontiers in Artificial Intelligence

Received: 13 April 2021

Accepted: 21 June 2021

Published: 16 July 2021

Citation:

Nishio M, Fujimoto K, Matsuo H,
Muramatsu C, Sakamoto R and
Fujita H (2021) Lung Cancer
Segmentation With Transfer Learning:
Usefulness of a Pretrained Model
Constructed From an Artificial Dataset
Generated Using a Generative
Adversarial Network.
Front. Artif. Intell. 4:694815.
doi: 10.3389/frai.2021.694815

Purpose: The purpose of this study was to develop and evaluate lung cancer segmentation with a pretrained model and transfer learning. The pretrained model was constructed from an artificial dataset generated using a generative adversarial network (GAN).

Materials and Methods: Three public datasets containing images of lung nodules/lung cancers were used: LUNA16 dataset, Decathlon lung dataset, and NSCLC radiogenomics. The LUNA16 dataset was used to generate an artificial dataset for lung cancer segmentation with the help of the GAN and 3D graph cut. Pretrained models were then constructed from the artificial dataset. Subsequently, the main segmentation model was constructed from the pretrained models and the Decathlon lung dataset. Finally, the NSCLC radiogenomics dataset was used to evaluate the main segmentation model. The Dice similarity coefficient (DSC) was used as a metric to evaluate the segmentation performance.

Results: The mean DSC for the NSCLC radiogenomics dataset improved overall when using the pretrained models. At maximum, the mean DSC was 0.09 higher with the pretrained model than that without it.

Conclusion: The proposed method comprising an artificial dataset and a pretrained model can improve lung cancer segmentation as confirmed in terms of the DSC metric. Moreover, the construction of the artificial dataset for the segmentation using the GAN and 3D graph cut was found to be feasible.

Keywords: lung cancer, lung nodule, segmentation, computed tomography, deep learning, generative adversarial network

INTRODUCTION

Segmentation of lung cancer is an important research topic, and various studies have been conducted so far. Segmentation results are used to determine the effectiveness of anticancer drugs (Mozley et al., 2012; Hayes et al., 2016) and to perform texture analyses on medical images (Bashir et al., 2017; Yang et al., 2020). To use the segmentation results of lung cancer effectively, the segmentation accuracy is an important factor. Segmentation is typically done manually by radiologists; however, manual segmentation can sometimes yield inaccurate results because of interobserver variability. Semiautomatic segmentation has lower interobserver variability than manual segmentation (Pfaehler et al., 2020). To overcome this interobserver variability, an automatic segmentation of lung cancer is desirable.

Recent years have witnessed significant development in the application of deep learning to various domains, including in the area of segmentation. For example, deep learning has been applied to the automatic segmentation of organs, such as the lungs, liver, pancreas, uterus, and bones, and to the automatic segmentation of tumors in these organs, with good segmentation performance (Roth et al., 2015; Chlebus et al., 2018; Isensee et al., 2018; Chen et al., 2019; Gordienko et al., 2019; Kurata et al., 2019; Noguchi et al., 2020; Hodneland et al., 2021).

One of the problems in the application of deep learning is a dataset. Deep learning does not perform well when the dataset is small. In general, it is difficult to increase the size of datasets containing medical images compared with other domains. This is due to the high cost of acquiring medical images and the need to protect personal information. To this end, transfer learning with pretrained models (Shin et al., 2016; Tschandl et al., 2019), data augmentation (Zhang et al., 2017; Yun et al., 2019), and artificial generation of datasets using generative adversarial networks (GANs) (Muramatsu et al., 2020) have been developed.

The GAN was first proposed by Goodfellow et al. (2014). The recent improvements made to the GAN have made it possible to generate high-quality, high-resolution images. Various attempts have been made to apply the GAN to medical image processing. Several studies have shown that it is possible to generate CT images of lung nodules using the GAN (Jin et al., 2018; Han et al., 2019; Onishi et al., 2019; Yang et al., 2019; Yi et al., 2019; Armanious et al., 2020).

To overcome the small dataset problem for segmentation, we proposed to use deep learning models pretrained with an artificially generated dataset using the GAN. We hypothesized that transfer learning with the proposed pretrained models could improve the automatic segmentation accuracy when using the lung cancer dataset. In general, a segmentation model obtained through supervised learning requires an image and its label as the dataset. In our study, to generate a dataset for segmentation, we used the GAN for image generation and the 3D graph cut method for generating labels of the generated images. No manual task for labeling was required to generate the dataset for pretraining.

MATERIALS AND METHODS

Our study used anonymized data extracted from public databases. Therefore, institutional review board approval was waived in accordance with the regulations of our country. **Figure 1** shows the outline of the proposed method for the segmentation model.

Dataset

Three public datasets containing computed tomography (CT) images were used: LUNg Nodule Analysis 2016 (LUNA16) dataset, Decathlon lung dataset, and NSCLC radiogenomics. **Table 1** shows a summary of the three datasets.

The LUNA16 dataset includes 888 sets of 3D CT images (Grand-Challenges, 2016; Setio et al., 2017) constructed for lung nodule detection. Therefore, the original LUNA16 dataset is unsuitable for segmentation. A previous study used the LUNA16 dataset to generate images of lung nodules using the GAN (Nishio et al., 2020a). We used the same dataset and a GAN model to generate the dataset for segmentation. For image preprocessing, the voxel size of the 3D CT images in the LUNA16 dataset was changed (1 mm × 1 mm × 1 mm isotropic). To generate lung cancer-like nodules and their labels in the LUNA16 dataset, large true nodules are problematic because labels of true nodules are not available in the LUNA16. Therefore, sets of 3D CT images with small lung nodules (the size of each nodule was <6 mm) were selected. As a result, 165 sets of 3D CT images in the LUNA16 dataset were used to generate an artificial dataset for segmentation.

The Decathlon challenge (<http://medicaldecathlon.com/>) was held to provide a fully open source and comprehensive benchmark for general purpose algorithmic validation and testing, covering several segmentation tasks. Decathlon includes several segmentation datasets, from which the Decathlon lung dataset (Task06) was used as the training and validation sets for our study. The Decathlon lung dataset includes 63 sets of 3D CT images and their segmentation labels. To simulate the small dataset, 10 and 30 sets of 3D CT images were selected from the Decathlon lung dataset; the image files of Decathlon lung dataset (NIfTI files) were sorted by file name, and the first 10 or 30 files were selected. As a result, three types of training datasets were prepared from the Decathlon lung dataset: 63 sets from the original Decathlon lung dataset (Decathlon_{full}), 30 sets (Decathlon_{mid}), and 10 sets (Decathlon_{small}). No image preprocessing was performed on the Decathlon lung dataset.

The NSCLC radiogenomics dataset (<https://wiki.cancerimagingarchive.net/display/Public/NSCLC-Radiomics>) contains images from 211 patients with non-small-cell lung cancer (Cancer Imaging Archive, 2021; Bakr et al., 2018; Gevaert et al., 2012; Clark et al., 2013). The dataset comprises CT, positron emission tomography/CT images, and segmentation maps of tumors in the CT scans. From the 211 patients, 3D CT images of 144 patients and their segmentation labels were selected for the current study. Segmentation labels are not available for the other 67 patients. The NSCLC radiogenomics dataset was used as the test set. For image preprocessing, the voxel size of the 3D CT images in the NSCLC radiogenomics dataset was changed (1 mm × 1 mm × 1 mm isotropic). The median volume of the lung

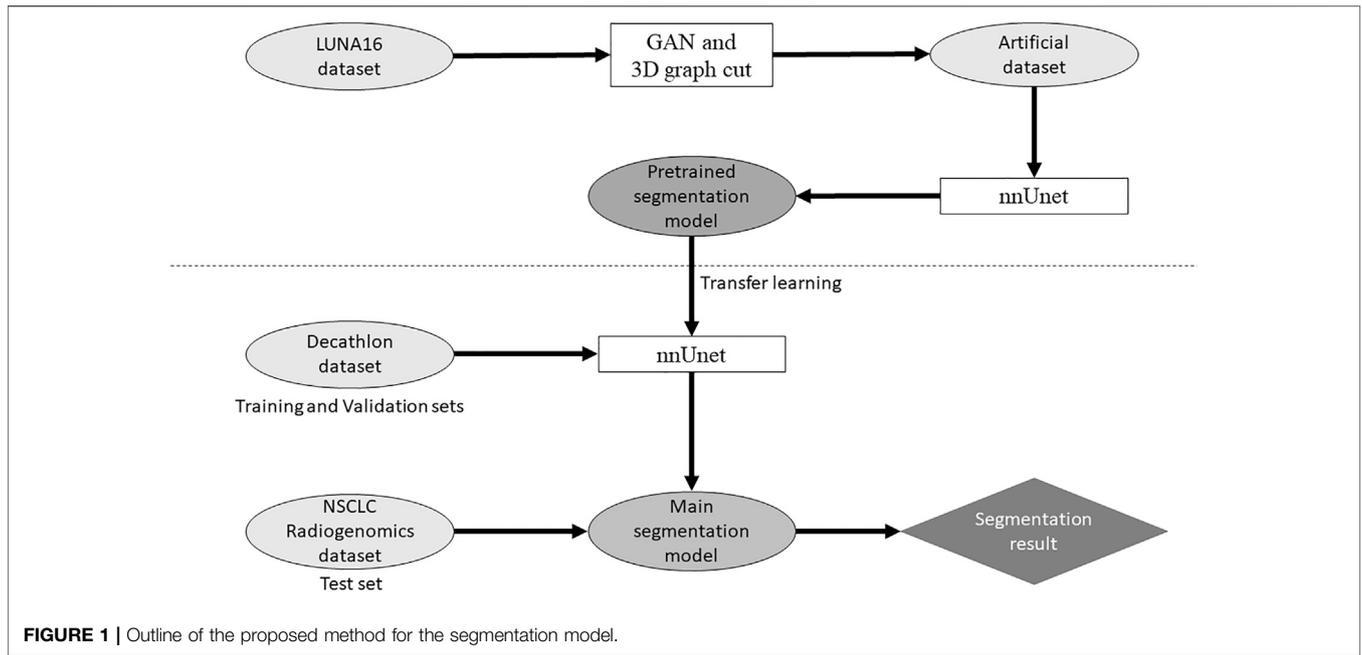


TABLE 1 | Summary of three datasets.

Dataset	Number of all CT scans	Number of CT scans used in this study	Selection criteria	Usage of dataset
LUNA16	888	165	3D CT images with small lung nodules (the size of each nodule was <6 mm) were selected	From the LUNA16 dataset, the GAN generated an artificial dataset with generated lung nodules. The pretrained model for lung cancer segmentation was obtained from the artificial dataset with original nnUnet.
Decathlon (task06, lung)	63	<ul style="list-style-type: none"> ■ 63 (Decathlon_{full}) ■ 30 (Decathlon_{mid}) ■ 10 (Decathlon_{small}) 	For Decathlon _{full} , no selection criteria. The image files of the Decathlon dataset were sorted by file name, and the first 10 or 30 files were selected for Decathlon _{small} and Decathlon _{mid}	Decathlon dataset was used for training and validation sets of lung cancer segmentation. Modified nnUnet was trained based on the validation set of the Decathlon dataset and the pretrained model.
NSCLC radiogenomics	211	144	Segmentation labels are available	NSCLC radiogenomics was used for test set of lung cancer segmentation (not used for the validation set).

Abbreviations: GAN, generative adversarial network.

cancer was 8,219 mm³ (interquartile range: 3,461.5–25,357 mm³) in the NSCLC radiogenomics dataset.

Dataset Generation

The LUNA16 dataset was used to generate an artificial dataset for segmentation. First, lung segmentation was performed for the chest CT images of the LUNA16 dataset, covering the lungs entirely. A pretrained deep learning model (a variant of U-net (Ronneberger et al., 2015)) was used for the lung segmentation (<https://github.com/JoHof/lungmask> (Hofmanninger et al., 2020)). Subsequently, 3D images of the nodule were generated using the GAN model, which is based on the variant of 3D pix2pix (Nishio et al., 2020a). While the GAN model can generate lung nodules at any location in the lungs, we used locations of true

nodules for nodule generation. In addition, we generated only one nodule per CT scan. To determine the location of the generated lung nodule, one location of true nodule was selected from the annotation of the LUNA16 dataset, for each CT scan. Therefore, the locations of generated lung nodules were fixed (no randomness). The true nodule was replaced with the nodule generated using the GAN model. For the nodule generation, 3D CT images were cropped with a volume of interest of 40 × 40 × 40 voxels for the location of the true nodules, and the cropped images were fed to the GAN model. While the size of the generated lung nodules can be adjusted with the GAN model, the GAN model generated the largest nodule as the model (the generation target size was 3 cm or higher). After nodule generation, the segmentation label was automatically generated

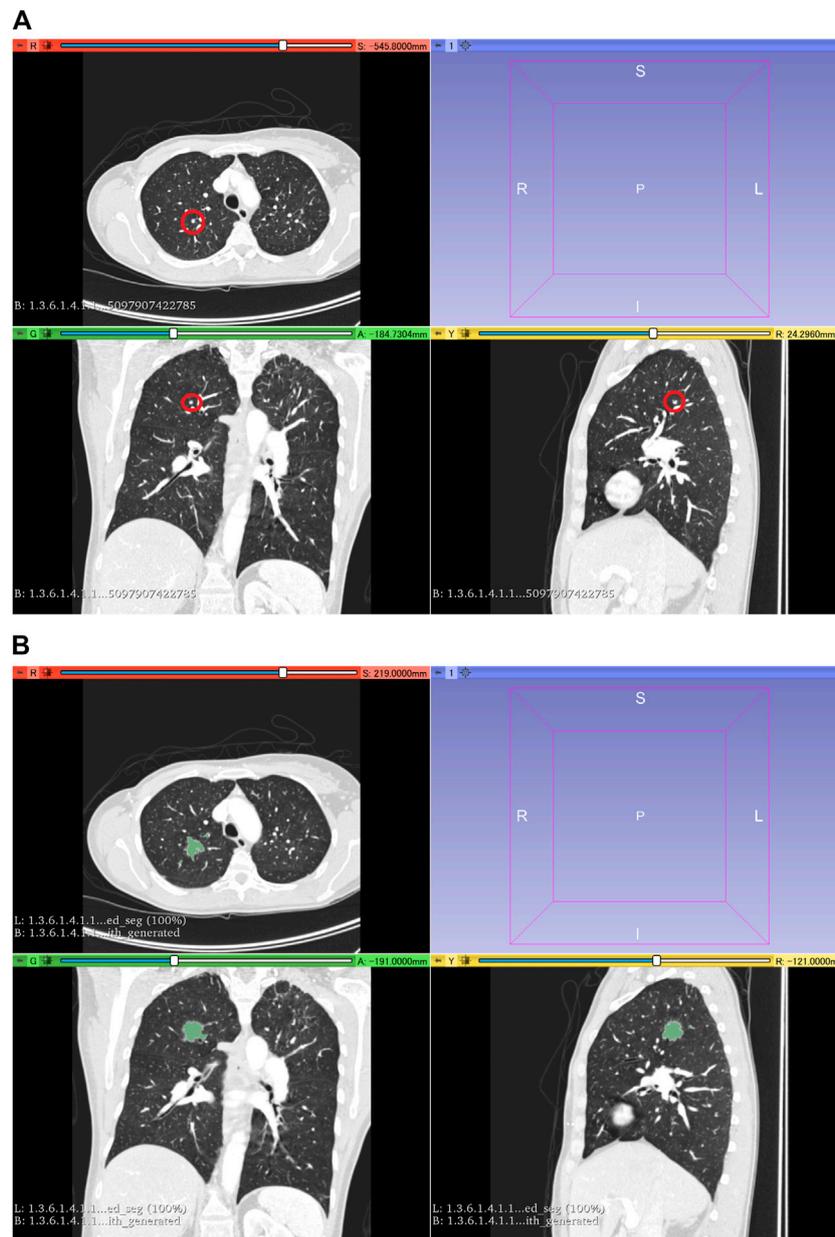


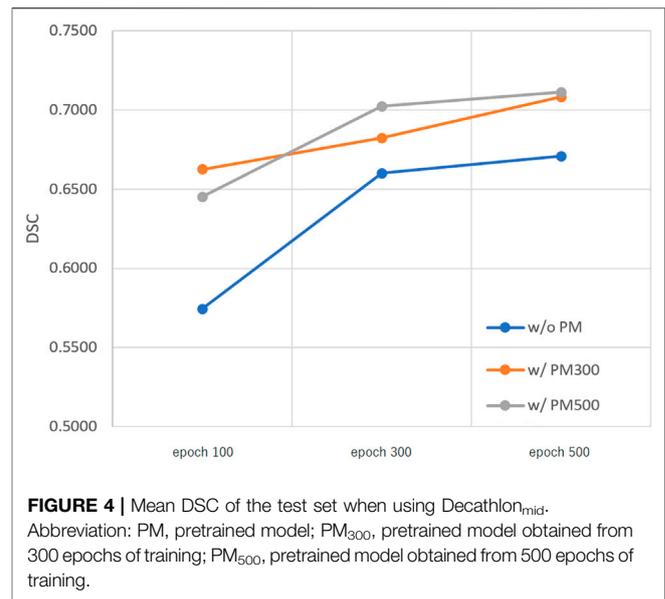
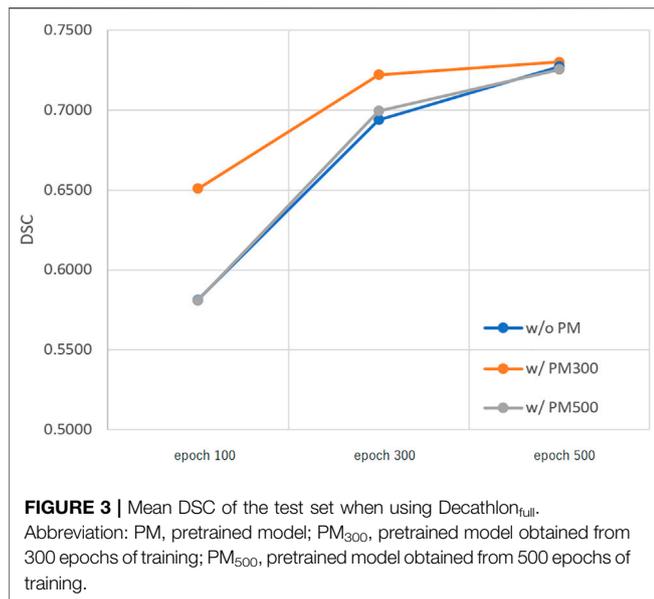
FIGURE 2 | 3D CT images of the chest. **(A)** Original CT images in the LUNA16 dataset. The red circle represents the true nodule specified in the LUNA16 dataset. **(B)** Lung nodule is artificially generated at the location of the true nodule. Label obtained with the 3D graph cut is superimposed on the 3D CT images.

using the 3D graph cut and Gaussian mixture (<https://github.com/mjirik/imcut>) (Jirik et al., 2013). Because the intensity of the seed point on the CT images was used to train the Gaussian mixture model, the center area of the generated images ($40 \times 40 \times 40$ voxels) was specified as seed points of the nodule, and the marginal area of the generated images was specified as seed points of the non-nodule (background). The output of the 3D graph cut was used as the segmentation label of the generated nodule. Next, the generated CT images of the nodule were merged with the original CT images. When merging the CT images of the generated nodules, only the areas that were assigned as lung

labels in the lung segmentation were targeted for the merging. The areas of the generated CT images that were assigned as non-lung labels were not merged. **Figure 2** shows the representative images of the generated nodules and their labels. In total, 165 lung nodules were generated for the 165 sets of 3D CT images in the LUNA16 dataset.

Segmentation Model

Open-source software (nnUnet) (Isensee et al., 2018) was used for the deep learning model of lung cancer segmentation, which is available at <https://github.com/MIC-DKFZ/nnUNet>. nnUnet is a



variant of U-net (Ronneberger et al., 2015). Originally, nnUnet was used for the Decathlon datasets (Isensee et al., 2018). Because the original nnUnet has no functionality of transfer learning, we modified the source code of nnUnet. With the modification, nnUnet could use a pretrained model and perform transfer learning. In addition, the number of epochs in the training nnUnet could be changed. Except for these two points, no modifications were made to nnUnet. Dataset splitting (training and validation sets) was performed with the default setting of nnUnet.

First, the generated dataset for segmentation obtained from the LUNA16 dataset was used to construct the pretrained model. Two pretrained models were built: one obtained from 300 epochs of training (PM₃₀₀) and the other obtained from 500 epochs of training (PM₅₀₀). Next, transfer learning using the two pretrained models was performed for the three Decathlon lung datasets (Decathlon_{full}, Decathlon_{mid}, and Decathlon_{small}) using the modified nnUnet. At this stage, no new layer was added to the model. Although several studies used layer freezing in transfer learning (Nishio et al., 2020b), no layers of the pretrained model were frozen in this study. To evaluate the effect of transfer learning, models were constructed without transfer learning (original nnUnet). Here, “original nnUnet” means that the source code of nnUnet was not changed, except for changing the number of epochs. The original nnUnet and its default setting were used for the model construction without transfer learning. In the model training, the epochs were set to 100, 300, and 500. The training of each model was started from epoch 1.

Evaluation of Segmentation Models

As the test dataset, 144 sets of 3D CT images from the NSCLC radiogenomics dataset were used. For each set, the Dice similarity coefficient (DSC) was used to evaluate the segmentation models. In addition, the Jaccard index (JI), sensitivity (SE), and specificity

(SP) were calculated as the evaluation metrics, which is expressed as follows:

$$DSC = \frac{2|P \cap L|}{|P| + |L|}, \quad (1)$$

$$JI = \frac{|P \cap L|}{|P| + |L| - |P \cap L|}, \quad (2)$$

$$SE = \frac{|P \cap L|}{|L|}, \quad (3)$$

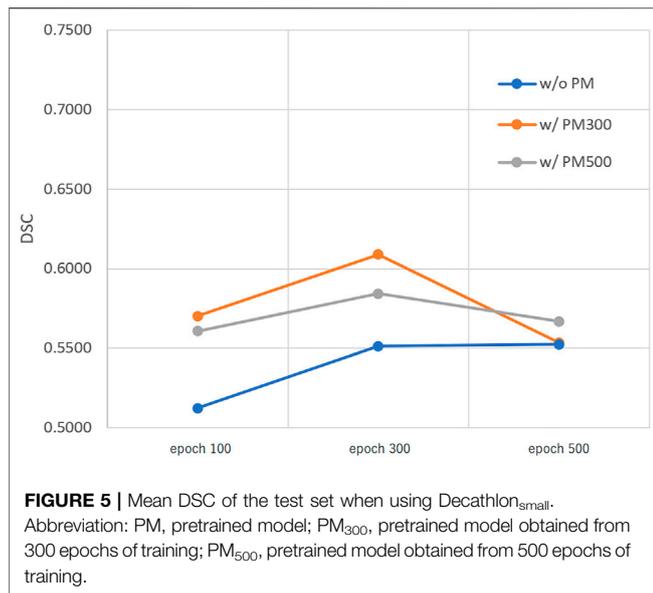
$$SP = \frac{|I| - (|L| + |P| - |P \cap L|)}{|I| - |L|}, \quad (4)$$

where $|P|$, $|L|$, and $|I|$ denote the number of voxels for the segmentation results, label of the lung cancer segmentation, and 3D CT images, respectively. $|P \cap L|$ represents the number of voxels where nnUnet can accurately segment the lung cancer (true positive). Before calculating the four metrics, a threshold of 0.5 was used to obtain the segmentation mask from the output of nnUnet.

Differences of DSC were statistically tested with the Wilcoxon signed rank test. To control the family-wise error rate, the Bonferroni correction was used; p -values less than 0.01666 were considered statistically significant. Statistical analyses were performed using R (version 4.0.4, <https://www.r-project.org/>).

RESULTS

Figures 3–5 show the mean DSC of the test set with and without PM₃₀₀ and PM₅₀₀ when Decathlon_{full}, Decathlon_{mid}, and Decathlon_{small} are used as the training sets, respectively. In these figures, the results without PM correspond to those of original nnUnet. Generally, PM₃₀₀ and PM₅₀₀ improved the mean DSC of nnUnet, compared with the original nnUnet (without the



pretrained model). In particular, the effectiveness of the pretrained model was high when using Decathlon_{mid} as the training set. Neither PM₃₀₀ nor PM₅₀₀ was useful for DSC improvement when Decathlon_{full} and Decathlon_{small} were used in the 500-epoch training. The DSC improvement was greater in the 100 and 300 epochs than that in the 500 epochs.

Tables 2–4 list the mean and standard deviation of the four metrics of the test set with and without PM₃₀₀ and PM₅₀₀ when Decathlon_{full}, Decathlon_{mid}, and Decathlon_{small} are used as the training sets, respectively. Because the volume ratio between cancer and noncancerous regions is extremely low, SP was extremely high in the current study. Regarding DSC, JI, and SE, the same trend can be observed. PM₃₀₀ and PM₅₀₀ improved the mean values of the three metrics; improvement in JI and SE was greater in the 100 and 300 epochs than that in the 500 epochs. Table 5 shows *p*-values for differences of DSC in Decathlon_{full}, Decathlon_{mid}, and Decathlon_{small}.

Figure 6 shows all the DSC values of the test set when using Decathlon_{mid} with and without the pretrained model. Figures 7, 8 show the representative segmentation results. Figures 7, 8 show the CT images in which PM is ineffective and effective, respectively. Supplementary Table 1 includes the segmentation results when the generated dataset consisted of variable-size-generated nodules. In addition, the Supplementary Table 2 includes visual evaluation results of cases with low DSC values.

DISCUSSION

In this study, we proposed a pretrained model for segmentation constructed from an artificial dataset of lung nodules generated using the GAN and 3D graph cut. Our results show that the

TABLE 2 | Segmentation metrics of test set using Decathlon_{full}.

Model	DSC mean	SD	JI mean	SD	SE mean	SD	SP mean	SD
W/o PM, epoch 100	0.5813	0.2495	0.4505	0.2352	0.7049	0.2514	0.99984	0.00027
W/PM ₃₀₀ , epoch 100	0.6510	0.2104	0.5150	0.2110	0.7672	0.1947	0.99987	0.00024
W/PM ₅₀₀ , epoch 100	0.5810	0.2474	0.4510	0.2411	0.7511	0.2133	0.99982	0.00020
W/o PM, epoch 300	0.6939	0.2409	0.5741	0.2360	0.7371	0.2542	0.99992	0.00011
W/PM ₃₀₀ , epoch 300	0.7221	0.2055	0.5995	0.2172	0.7580	0.2102	0.99991	0.00023
W/PM ₅₀₀ , epoch 300	0.6995	0.2302	0.5789	0.2347	0.7514	0.2119	0.99992	0.00012
W/o PM, epoch 500	0.7273	0.2266	0.6120	0.2315	0.7334	0.2464	0.99995	0.00009
W/PM ₃₀₀ , epoch 500	0.7301	0.2145	0.6122	0.2239	0.7662	0.2061	0.99990	0.00025
W/PM ₅₀₀ , epoch 500	0.7255	0.2424	0.6139	0.2386	0.7428	0.2484	0.99993	0.00012

Abbreviations: DSC, Dice similarity coefficients; JI, Jaccard index; SE, sensitivity; SP, specificity; PM, pretrained model; PM₃₀₀, pretrained model obtained from 300 epochs of training; PM₅₀₀, pretrained model obtained from 500 epochs of training.

TABLE 3 | Segmentation metrics of test set using Decathlon_{mid}.

Model	DSC mean	SD	JI mean	SD	SE mean	SD	SP mean	SD
W/o PM, epoch 100	0.5744	0.2704	0.4493	0.2486	0.6122	0.2830	0.99993	0.00008
W/PM ₃₀₀ , epoch 100	0.6624	0.2305	0.5342	0.2303	0.6999	0.2221	0.99992	0.00012
W/PM ₅₀₀ , epoch 100	0.6452	0.2351	0.5155	0.2302	0.7144	0.2302	0.99990	0.00015
W/o PM, epoch 300	0.6600	0.2572	0.5398	0.2509	0.6750	0.2579	0.99993	0.00013
W/PM ₃₀₀ , epoch 300	0.6822	0.2470	0.5631	0.2463	0.7234	0.2294	0.99991	0.00015
W/PM ₅₀₀ , epoch 300	0.7024	0.2308	0.5823	0.2344	0.7292	0.2201	0.99990	0.00022
W/o PM, epoch 500	0.6708	0.2689	0.5570	0.2629	0.6892	0.2608	0.99991	0.00019
W/PM ₃₀₀ , epoch 500	0.7083	0.2039	0.5819	0.2160	0.7479	0.1936	0.99992	0.00012
W/PM ₅₀₀ , epoch 500	0.7112	0.2228	0.5909	0.2293	0.7266	0.2215	0.99993	0.00013

Abbreviations: DSC, Dice similarity coefficients; JI, Jaccard index; SE, sensitivity; SP, specificity; PM, pretrained model; PM₃₀₀, pretrained model obtained from 300 epochs of training; PM₅₀₀, pretrained model obtained from 500 epochs of training.

TABLE 4 | Segmentation metrics of test set using Decathlon_{small}.

Model	DSC mean	SD	JI mean	SD	SE mean	SD	SP mean	SD
W/o PM, epoch 100	0.5124	0.2857	0.3928	0.2555	0.4671	0.2838	0.99997	0.00007
W/PM ₃₀₀ , epoch 100	0.5702	0.2513	0.4400	0.2388	0.5183	0.2524	0.99996	0.00007
W/PM ₅₀₀ , epoch 100	0.5608	0.2681	0.4352	0.2491	0.5190	0.2761	0.99997	0.00006
W/o PM, epoch 300	0.5515	0.2742	0.4274	0.2510	0.4950	0.2811	0.99997	0.00007
W/PM ₃₀₀ , epoch 300	0.6090	0.2414	0.4781	0.2361	0.5511	0.2510	0.99997	0.00007
W/PM ₅₀₀ , epoch 300	0.5844	0.2607	0.4574	0.2466	0.5215	0.2719	0.99997	0.00006
W/o PM, epoch 500	0.5525	0.2855	0.4316	0.2573	0.5109	0.2851	0.99997	0.00006
W/PM ₃₀₀ , epoch 500	0.5536	0.2856	0.4338	0.2629	0.4934	0.2878	0.99998	0.00005
W/PM ₅₀₀ , epoch 500	0.5670	0.2731	0.4435	0.2557	0.5089	0.2786	0.99997	0.00006

Abbreviations: DSC, Dice similarity coefficients; JI, Jaccard index; SE, sensitivity; SP, specificity; PM, pretrained model; PM₃₀₀, pretrained model obtained from 300 epochs of training; PM₅₀₀, pretrained model obtained from 500 epochs of training.

accuracy of lung cancer segmentation could be improved when this pretrained model was used for transfer learning in the segmentation process. The effectiveness of the pretrained model was higher on the Decathlon_{mid} and Decathlon_{small} datasets than that of the pretrained model on the Decathlon_{full} dataset, suggesting that our proposed method may be effective on small datasets.

The pretrained model was more effective when the number of training epochs was low. In other words, the number of epochs required to achieve a sufficient segmentation performance was

lower with the pretrained model than without it. This may be attributed to the fact that the pretrained model provides good initial values for the trainable parameters of nnUnet.

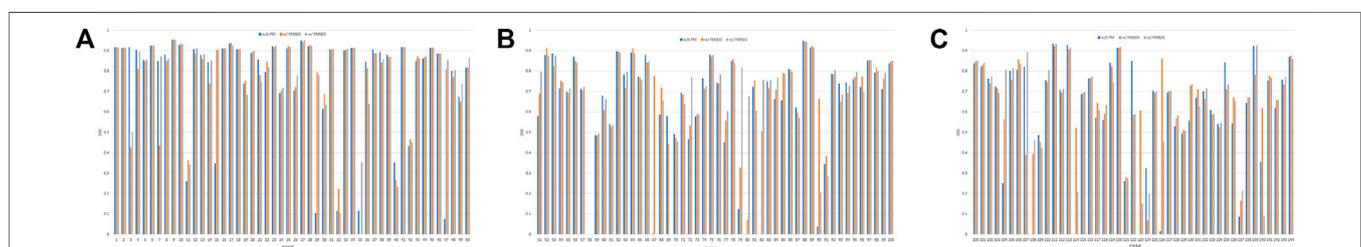
Previously, a study used U-net and GAN combinedly for multi-organ segmentation on 3D CT images (Dong et al., 2019). However, the study did not use a pretrained model. Another study was conducted on a classification model using a dataset generated with GANs and a pretrained model (Onishi et al., 2020). To the best of our knowledge, no studies have been reported on segmentation models with GANs and a pretrained model. Our results and those of Onishi et al. (2020) indicate that the GAN generated dataset, and its pretrained models may be useful for various tasks.

Several studies have reported the use of artificially generated datasets using the GAN for data augmentation (Jin et al., 2018; Onishi et al., 2019; Yang et al., 2019; Muramatsu et al., 2020). Similarly, in this study, we tried to use a dataset generated using the GAN for data augmentation. However, we could not obtain effective results for lung cancer segmentation when the artificial dataset was used as data augmentation (data not shown in this article). Instead, we constructed a pretrained model for the segmentation using the generated lung nodules and performed transfer learning based on the pretrained model, yielding higher lung cancer segmentation accuracy. Although it was difficult to perform accurate classification between the generated lung nodules and the true lung nodules (Nishio et al., 2020a), the generated lung nodules had little variation as lung cancer. It is speculated that mixing the generated lung nodules with the true lung nodules could distort the distribution as the dataset of lung

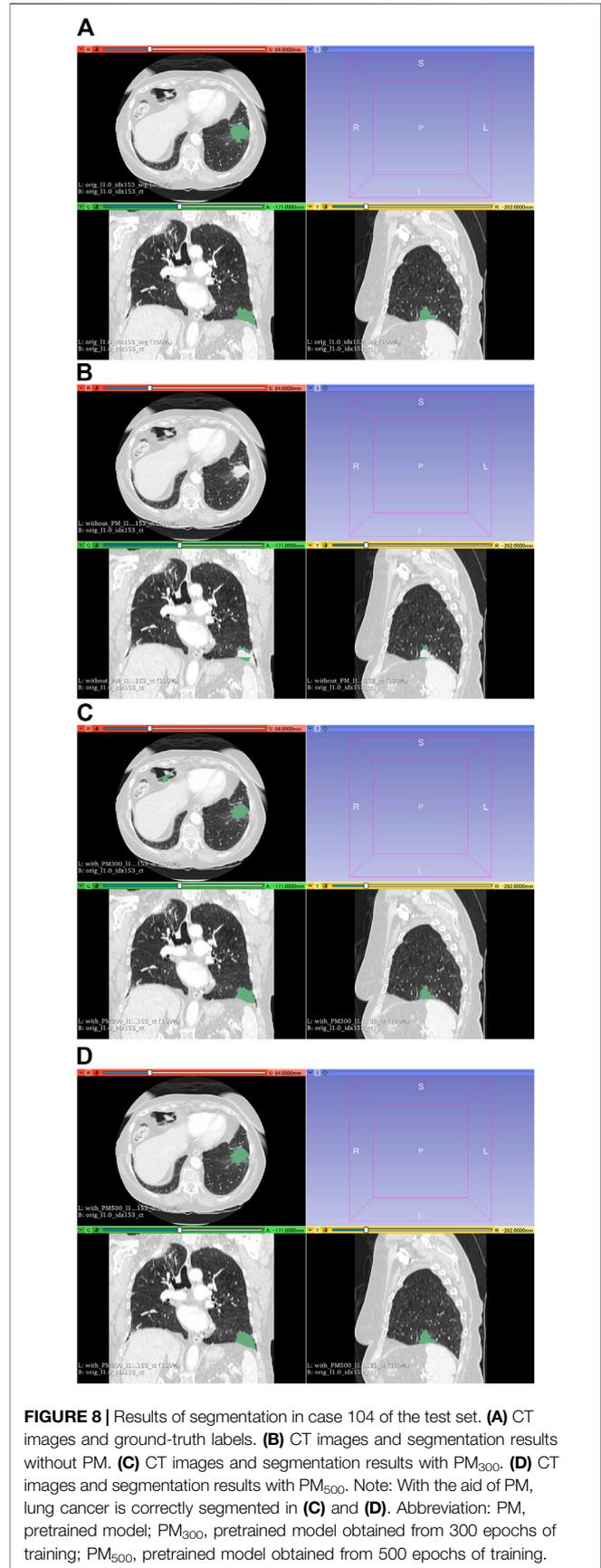
TABLE 5 | p -values for differences of DSC in Decathlon_{full}, Decathlon_{mid}, and Decathlon_{small}.

Pair	Decathlon _{full}	Decathlon _{mid}	Decathlon _{small}
Epoch 100	—	—	—
w/o PM vs. w/PM ₃₀₀	7.857×10^{-5}	1.877×10^{-7}	0.0006931
w/o PM vs. w/PM ₅₀₀	0.9682	0.0001075	0.0009121
w/o PM ₃₀₀ vs. w/PM ₅₀₀	1.606×10^{-5}	0.4642	0.1499
Epoch 300	—	—	—
w/o PM vs. w/PM ₃₀₀	0.2346	0.04381	6.662×10^{-5}
w/o PM vs. w/PM ₅₀₀	0.7664	0.000649	0.001194
w/o PM ₃₀₀ vs. w/PM ₅₀₀	0.1493	0.01325	0.3050
Epoch 500	—	—	—
w/o PM vs. w/PM ₃₀₀	0.9743	0.2322	0.5990
w/o PM vs. w/PM ₅₀₀	0.7243	0.005435	0.8746
w/o PM ₃₀₀ vs. w/PM ₅₀₀	0.9534	0.06595	0.9229

Abbreviations: DSC, Dice similarity coefficients; PM, pretrained model; PM₃₀₀, pretrained model obtained from 300 epochs of training; PM₅₀₀, pretrained model obtained from 500 epochs of training.

**FIGURE 6** | DSC values of the test set when using Decathlon_{mid} with and without the pretrained model. (A) Cases 1–50, (B) cases 51–100, and (C) cases 101–144.

Note: DSC values are obtained with models obtained from 500 epochs of training. Abbreviation: PM, pretrained model; PM₃₀₀, pretrained model obtained from 300 epochs of training; PM₅₀₀, pretrained model obtained from 500 epochs of training.



cancer segmentation and adversely affect the model training of nnUnet.

Generally, supervised learning (e.g., nnUnet) requires annotation data as the dataset. On the datasets of lung cancer segmentation, clinicians frequently annotate lung cancer on CT images to build lung cancer datasets, which is time consuming and labor intensive. Although it is possible to manually annotate the generated data of our dataset, we decided to use the 3D graph cut to obtain annotation data of the generated lung nodules. This made it possible to build an artificial dataset for the segmentation without requiring any manual task.

Although the generated lung nodules and the pretrained model based on them could effectively improve the accuracy of lung cancer segmentation, this pretrained model is not always effective. For example, the effectiveness of the pretrained model was not observed in the 500-epoch training of Decathlon_{full} and Decathlon_{small}. For the former case, this was attributed to the fact that Decathlon_{full} had sufficient amount of data and the number of training epochs was high. In the latter, the number of datasets was very small (10 cases). Therefore, even when the pretrained model was used, the training segmentation model was unstable, and the effectiveness of the pretrained model was limited.

Our study has some limitations. First, we used three public datasets containing images of lung nodules and/or lung cancer. However, we did not verify whether the generalizability of our segmentation model can be improved under external variation. Second, we focused on lung nodules and/or lung cancer in the current study. Therefore, the effectiveness of our method for other diseases or other organs has not been validated. In particular, it is necessary to confirm whether the automatic generation of annotation data using the 3D graph cut can be applied to other diseases and other organs. Third, because of the GAN model's limitation (Nishio et al., 2020a), it was impossible to generate lung nodules larger than 40 mm. Therefore, the effect of large generated nodules is not investigated in the current study.

In conclusion, the proposed method comprising an artificial dataset and a pretrained model can improve the accuracy of lung cancer segmentation; however, it should be further investigated for other diseases and other organs.

REFERENCES

- Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., et al. (2020). MedGAN: Medical Image Translation Using GANs. *Comput. Med. Imaging Graphics* 79, 101684. doi:10.1016/j.compmedimag.2019.101684
- Bakr, S., Gevaert, O., Echeqaray, S., Ayers, K., Zhou, M., Shafiq, M., et al. (2018). A Radiogenomic Dataset of Non-small Cell Lung Cancer. *Sci. Data* 5, 1–9. doi:10.1038/sdata.2018.202
- Bashir, U., Azad, G., Siddique, M. M., Dhillon, S., Patel, N., Bassett, P., et al. (2017). The Effects of Segmentation Algorithms on the Measurement of 18F-FDG PET Texture Parameters in Non-small Cell Lung Cancer. *EJNMMI Res.* 7, 60. doi:10.1186/s13550-017-0310-3
- Chen, W., Wei, H., Peng, S., Sun, J., Qiao, X., and Liu, B. (2019). HSN: Hybrid Segmentation Network for Small Cell Lung Cancer Segmentation. *IEEE Access* 7, 75591–75603. doi:10.1109/ACCESS.2019.2921434
- Chlebus, G., Schenk, A., Moltz, J. H., van Ginneken, B., Hahn, H. K., and Meine, H. (2018). Automatic Liver Tumor Segmentation in CT with Fully Convolutional

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: 1. <https://luna16.grand-challenge.org/> 2. <http://medicaldecathlon.com/> 3. <https://wiki.cancerimagingarchive.net/display/Public/NSCLC-Radiomics>.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MN contributed to conception and design of the study. MN and KF organized the database. MN and HM developed the software. MN performed the statistical analysis. MN wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was supported by JSPS KAKENHI (grant numbers: 19H03599 and JP19K17232). The funder had no role in the present study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frai.2021.694815/full#supplementary-material>

Neural Networks and Object-Based Postprocessing. *Sci. Rep.* 8, 1–7. doi:10.1038/s41598-018-33860-7

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., et al. (2013). The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging* 26, 1045–1057. doi:10.1007/s10278-013-9622-7

Dong, X., Lei, Y., Wang, T., Thomas, M., Tang, L., Curran, W. J., et al. (2019). Automatic Multiorgan Segmentation in thoraxCTimages Using U-net-GAN. *Med. Phys.* 46, 2157–2168. doi:10.1002/mp.13458

Gevaert, O., Xu, J., Hoang, C. D., Leung, A. N., Xu, Y., Quon, A., et al. (2012). Non-Small Cell Lung Cancer: Identifying Prognostic Imaging Biomarkers by Leveraging Public Gene Expression Microarray Data-Methods and Preliminary Results. *Radiology* 264, 387–396. doi:10.1148/radiol.12111607

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). “Generative Adversarial Nets.” in Proceedings of the Advances in Neural Information Processing Systems. Neural information processing systems foundation. pp. 2672–2680.

Gordienko, Y., Gang, P., Hui, J., Zeng, W., Kochura, Y., Alienin, O., et al. (2019). “Deep Learning with Lung Segmentation and Bone Shadow Exclusion

- Techniques for Chest X-ray Analysis of Lung Cancer,” in *Advances in Intelligent Systems and Computing* (Berlin, Germany: Springer-Verlag), 638–647. doi:10.1007/978-3-319-91008-6_63
- Ginneken, B. V., Kerkstra, S., and Meakin, J. (2016). LUNg Nodule Analysis. *Open Med. Image Comput. [Internet]*. Available at: <https://luna16.grand-challenge.org/>
- Han, C., Kitamura, Y., Kudo, A., Ichinose, A., Rundo, L., Furukawa, Y., et al. (2019). “Synthesizing Diverse Lung Nodules Wherever Massively: 3D Multi-Conditional GAN-based CT Image Augmentation for Object Detection,” in *Proceedings - Int Conf 3D Vision*. 3DV, 729–737. Available at: <http://arxiv.org/abs/1906.04962>
- Hayes, S. A., Pietanza, M. C., O’Driscoll, D., Zheng, J., Moskowitz, C. S., Kris, M. G., et al. (2016). Comparison of CT Volumetric Measurement with RECIST Response in Patients with Lung Cancer. *Eur. J. Radiol.* 85, 524–533. doi:10.1016/j.ejrad.2015.12.019
- Hodneland, E., Dybvik, J. A., Wagner-Larsen, K. S., Šoltészová, V., Munthe-Kaas, A. Z., Fasmer, K. E., et al. (2021). Automated Segmentation of Endometrial Cancer on MR Images Using Deep Learning. *Sci. Rep.* 11, 179. doi:10.1038/s41598-020-80068-9
- Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., and Langs, G. (2020). Automatic Lung Segmentation in Routine Imaging Is Primarily a Data Diversity Problem, Not a Methodology Problem. *Eur. Radiol. Exp.* 4, 1–13. doi:10.1186/s41747-020-00173-2
- Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., et al. (2018). “NNU-NET: Self-Adapting Framework for U-Net-Based Medical Image Segmentation,” in *MICCAI 2018, 21st International Conference, Granada, Spain, September 16–20, 2018*. Available at: <http://arxiv.org/abs/1809.10486> (Accessed 5 Apr 2021)
- Jin, D., Xu, Z., Tang, Y., Harrison, A. P., and Mollura, D. J. (2018). CT-realistic Lung Nodule Simulation from ds Conditional Generative Adversarial Networks for Robust Lung Segmentation. *Lect Notes Computer Science (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. ;11071, 732–740. Available at: <http://arxiv.org/abs/1806.04051>
- Jirik, M., Lukes, V., Svobodová, M., and Zelezný, M. (2013). “Image Segmentation in Medical Imaging via Graph-Cuts,” in *11th International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-11-2013)*. Samara, Russia, September 2013.
- Kurata, Y., Nishio, M., Kido, A., Fujimoto, K., Yakami, M., Isoda, H., et al. (2019). Automatic Segmentation of the Uterus on MRI Using a Convolutional Neural Network. *Comput. Biol. Med.* 114, 103438. doi:10.1016/j.combiomed.2019.103438
- Mozley, P. D., Bendtsen, C., Zhao, B., Schwartz, L. H., Thorn, M., Rong, Y., et al. (2012). Measurement of Tumor Volumes Improves RECIST-Based Response Assessments in Advanced Lung Cancer. *Translational Oncol.* 5, 19–25. doi:10.1593/tlo.11232
- Muramatsu, C., Nishio, M., Goto, T., Oiwa, M., Morita, T., Yakami, M., et al. (2020). Improving Breast Mass Classification by Shared Data with Domain Transformation Using a Generative Adversarial Network. *Comput. Biol. Med.* 119, 103698. doi:10.1016/j.combiomed.2020.103698
- Nishio, M., Muramatsu, C., Noguchi, S., Nakai, H., Fujimoto, K., Sakamoto, R., et al. (2020). Attribute-guided Image Generation of Three-Dimensional Computed Tomography Images of Lung Nodules Using a Generative Adversarial Network. *Comput. Biol. Med.* 126, 104032. doi:10.1016/j.combiomed.2020.104032
- Nishio, M., Noguchi, S., Matsuo, H., and Murakami, T. (2020). Automatic Classification between COVID-19 Pneumonia, Non-COVID-19 Pneumonia, and the Healthy on Chest X-ray Image: Combination of Data Augmentation Methods. *Sci. Rep.* 10, 1–6. doi:10.1038/s41598-020-74539-2
- Noguchi, S., Nishio, M., Yakami, M., Nakagomi, K., and Togashi, K. (2020). Bone Segmentation on Whole-Body CT Using Convolutional Neural Network with Novel Data Augmentation Techniques. *Comput. Biol. Med.* 121, 103767. doi:10.1016/j.combiomed.2020.103767
- Cancer Imaging Archive (2021). *NSCLC Radiogenomics - the Cancer Imaging Archive*. (TCIA) Public Access - Cancer Imaging Archive Wiki. Available at: <https://wiki.cancerimagingarchive.net/display/Public/NSCLC+Radiogenomics> (Accessed 5 Apr 2021)
- Onishi, Y., Teramoto, A., Tsujimoto, M., Tsukamoto, T., Saito, K., Toyama, H., et al. (2019). Automated Pulmonary Nodule Classification in Computed Tomography Images Using a Deep Convolutional Neural Network Trained by Generative Adversarial Networks. *Biomed. Res. Int.* 2019, 1–9. doi:10.1155/2019/6051939
- Onishi, Y., Teramoto, A., Tsujimoto, M., Tsukamoto, T., Saito, K., Toyama, H., et al. (2020). Multiplanar Analysis for Pulmonary Nodule Classification in CT Images Using Deep Convolutional Neural Network and Generative Adversarial Networks. *Int. J. CARS* 15, 173–178. doi:10.1007/s11548-019-02092-z
- Pfahler, E., Burggraaff, C., Kramer, G., Zijlstra, J., Hoekstra, O. S., Jalving, M., et al. (2020). “PET Segmentation of Bulky Tumors: Strategies and Workflows to Improve Inter-observer Variability. Zeng L, editor. *PLoS One* 15, e0230901. doi:10.1371/journal.pone.0230901
- Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-net: Convolutional Networks for Biomedical Image Segmentation,” in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Berlin, Germany: Springer-Verlag), 234–241. doi:10.1007/978-3-319-24574-4_28
- Roth, H. R., Lu, L., Farag, A., Shin, H.-C., Liu, J., Turkbey, E. B., et al. (2015). “Deeporgan: Multi-Level Deep Convolutional Networks for Automated Pancreas Segmentation,” in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Berlin, Germany: Springer-Verlag), 556–564. doi:10.1007/978-3-319-24553-9_68
- Setio, A. A. A., Traverso, A., de Bel, T., Berens, M. S. N., Bogaard, C. v. d., Cerello, P., et al. (2017). Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography Images: The LUNA16 challenge. *Med. Image Anal.* 42, 1, 13. doi:10.1016/j.media.2017.06.015
- Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* 35, 1285–1298. doi:10.1109/TMI.2016.2528162
- Tschandl, P., Sinz, C., and Kittler, H. (2019). Domain-specific Classification-Pretrained Fully Convolutional Network Encoders for Skin Lesion Segmentation. *Comput. Biol. Med.* 104, 111–116. doi:10.1016/j.combiomed.2018.11.010
- Yang, F., Simpson, G., Young, L., Ford, J., Dogan, N., and Wang, L. (2020). Impact of Contouring Variability on Oncological PET Radiomics Features in the Lung. *Sci. Rep.* 10, 1–10. doi:10.1038/s41598-019-57171-7
- Yang, J., Liu, S., Grbic, S., Setio, A. A. A., Xu, Z., Gibson, E., et al. (2019). “Class-aware Adversarial Lung Nodule Synthesis in Ct Images,” in *Proceedings - International Symposium on Biomedical Imaging*. IEEE Computer Society, 1348–1352. doi:10.1109/ISBI.2019.8759493
- Yi, X., Walia, E., and Babyn, P. (2019). Generative Adversarial Network in Medical Imaging: A Review. *Med. Image Anal.* 58, 101552. doi:10.1016/j.media.2019.101552
- Yun, S., Han, D., Chun, S., Oh, S. J., Yoo, Y., and Choe, J. (2019). “CutMix: Regularization Strategy to Train strong Classifiers with Localizable Features,” in *Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, November 2019* (. Institute of Electrical and Electronics Engineers Inc. (IEEE)), 6022–6031. doi:10.1109/ICCV.2019.00612
- Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. (2017). MIXUP: Beyond Empirical Risk Minimization. *arXiv*. Available at: <http://arxiv.org/abs/1710.09412> (Accessed April 5, 2021).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Nishio, Fujimoto, Matsuo, Muramatsu, Sakamoto and Fujita. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.