Journal of Digital Imaging (2023) 36:1237–1247
https://doi.org/10.1007/s10278-022-00772-y

Effect of Contrast Level and Image Format on a Deep Learning Algorithm for the Detection of Pneumothorax with Chest Radiography

Myeong Seong Yoon[1,3,7] · Gitaek Kwon[2,4] · Jaehoon Oh[1,3] · Jongbin Ryu[5] · Jongwoo Lim[2,3] · Bo-kyeong Kang[3,6] · Juncheol Lee[1] · Dong-Kyoon Han[7]

Received: 30 March 2022 / Revised: 23 December 2022 / Accepted: 29 December 2022 / Published online: 25 January 2023

Abstract

Under the black-box nature in the deep learning model, it is uncertain how the change in contrast level and format affects the performance. We aimed to investigate the effect of contrast level and image format on the effectiveness of deep learning for diagnosing pneumothorax on chest radiographs. We collected 3316 images (1016 pneumothorax and 2300 normal images), and all images were set to the standard contrast level (100%) and stored in the Digital Imaging and Communication in Medicine and Joint Photographic Experts Group (JPEG) formats. Data were randomly separated into 80% of training and 20% of test sets, and the contrast of images in the test set was changed to 5 levels (50%, 75%, 100%, 125%, and 150%). We trained the model to detect pneumothorax using ResNet-50 with 100% level images and tested with 5-level images in the two formats. While comparing the overall performance between each contrast level in the two formats, the area under the receiver-operating characteristic curve (AUC) was significantly different (all $p < 0.001$) except between 125 and 150% in JPEG format ($p = 0.382$). When comparing the two formats at same contrast levels, AUC was significantly different (all $p < 0.001$) except 50% and 100% ($p = 0.079$ and $p = 0.082$, respectively). The contrast level and format of medical images could influence the performance of the deep learning model. It is required to train with various contrast levels and formats of image, and further image processing for improvement and maintenance of the performance.

Keywords Deep learning · Pneumothorax · Image format · Contrast level · Artificial intelligence

Introduction

Owing to the remarkable development of deep learning, the importance of medical image recognition using machine learning has gradually increased [1–3]. Recent advances offer improvements in chest radiograph interpretation, and several machine learning models achieve radiologist- or radiology resident–level performance for various chest radiograph findings [4–7].

Pneumothorax is a thoracic emergency because the presence of air within the pleural space outside the lung causes collapse of the lung, resulting in respiratory failure [8].

Myeong Seong Yoon and Gitaek Kwon contributed to this work equally as first authors.

✉ Jaehoon Oh
ojjai@hanmail.net; ohjae7712@gmail.com

✉ Jongbin Ryu
jongbin.ryu@gmail.com

1 Department of Emergency Medicine, College of Medicine, Hanyang University, 222 Wangsimni-Ro, Seongdong-Gu, Seoul 04763, Republic of Korea

2 Department of Computer Science, Hanyang University, 222 Wangsimni-Ro, Seongdong-Gu, Seoul 04763, Republic of Korea

3 Machine Learning Research Center for Medical Data, Hanyang University, 222 Wangsimni-Ro, Seongdong-Gu, Seoul 04763, Republic of Korea

4 VUNO, Inc, 479 Gangnam-daero, Seocho-gu, Seoul 06541, Republic of Korea

5 Department of Software and Computer Engineering, Ajou University, 206 World cup-ro, Suwon-si, Gyeonggi Do 16499, Republic of Korea

6 Department of Radiology, College of Medicine, Hanyang University, 222 Wangsimni-Ro, Seongdong-Gu, Seoul 04763, Republic of Korea

7 Department of Radiological Science, Eulji University, 553 Sanseong-daero, Seongnam-si, Gyeonggi Do 13135, Republic of Korea

Chest radiography is a commonly used screening tool for pneumothorax detection. The sensitivity for pneumothorax detection on chest radiographs with clinicians has been reported to be approximately 80–86% [9, 10]. However, it sometimes fails to detect pneumothorax and could be too dependent on the clinician's experience, including shape, size, and patient positioning [11–13].

Recently, many deep learning studies for pneumothorax detection have been suggested to improve its performance and aid clinicians in finding it on chest radiographs [14–21]. However, the performance decreased when validated with an external dataset in some studies [18, 20]. These reasons might originate from the heterogeneity of images due to patient characteristics, radiographic device and radiation dose, contrast level, and format in external data [18, 20, 22]. Contrast level or signal-to-noise ratio could be a significant factor in performance degradation for deep learning models in general photography and computer vision fields [23–25]. The preprocessing technique could overcome a large part of these differences. However, the effectiveness can be different on the results of a network performance according to the preprocessing methods and not eliminate all systematic difference [26, 27]. To achieve a comparative performance on deep learning models, high-quality and well-curated training datasets should be prepared to solve the heterogeneity of images [28–30].

In the deep learning model, it is uncertain how the change in contrast level and format affects the performance. Previous studies did not consider contrast levels and image formats on chest radiographs for detecting pneumothorax [14–21]. This study hypothesized that the performance of the deep learning model would differ according to the contrast level and image format between the training and test datasets. Therefore, the aim of this study was to investigate how the contrast level and image format affect the performance of deep learning models for the detection of pneumothorax on chest radiographs. The remaining section of the paper is organized as follows. The "Methods" section presents the study design, the data collection, the data preprocessing, the experiment, the network of the proposed model, and the primary outcome. The "Results" section describes the experimental setup, the results of overall performance, and the comparison of outcomes. Finally, the "Discussion" section discusses our findings compared with the previous research.

This work could contribute for the deep learning research field with medical images and summarized as follows.

- The difference of contrast level on the training and test set images could influence the performance of the deep learning. We evaluated the network performance for the detection of pneumothorax on chest radiographs by five tests with 50%, 75%, 100, 125%, and 150% contrast levels after training with 100% standard contrast level and ResNet-50 model.
- 12- to 16-bit based the Digital Imaging and Communication in Medicine (DICOM) data are mainly used in real clinical situations whereas 8-bit image formats such as Portable Network Graphics (PNG) and Joint Photographic Experts Group (JPEG) are mainly used to train models in deep learning with medical image. The information loss DICOM data could occur when it is converted to an 8-bit image format which is favorable for file storage and transmission. We evaluated the performance of deep learning according to image formats in this experiment.
- This study showed that performances for the detection of pneumothorax were different according to image contrast levels and formats on chest radiographs. This research could confirm that the performance of deep learning could be influenced according to image contrast level and formats in the dataset.

Methods

Study Design

The overall flow of this study is illustrated in Fig. 1. This retrospective study was conducted at a tertiary academic hospital (Seoul, Republic of Korea) between May and November 2020. This study was authorized by the Institutional Review Board (IRB) of our university hospital (ref. no. HYUH 2020-03-039). The institutional review boards at our university hospital waived the necessity for informed consent. All methodologies and procedures were carried out in line with the Helsinki Declaration.

Data Collection

We gathered a list of patients who were 18–80 years old and visited with complaints of chest pain and dyspnea at the emergency department of our university hospital between the dates of January 2012 and September 2020 using medical records.

Images were obtained from chest radiographs of patients who were confirmed to have pneumothorax by radiologists using chest computed tomography scans among the above lists. We excluded images of patients who underwent lobectomy, those with severe deformity, massive hemothorax or mass, and chest tube. Normal images were conveniently sampled from chest radiographs, which were not indicative of pneumothorax, and classified as "unremarkable study," "non-specific finding," and "no active lesion" by radiologists. The collected images had a pneumothorax-to-normal ratio of approximately 1:2. All candidate images were extracted and stored in the DICOM format using the Picture

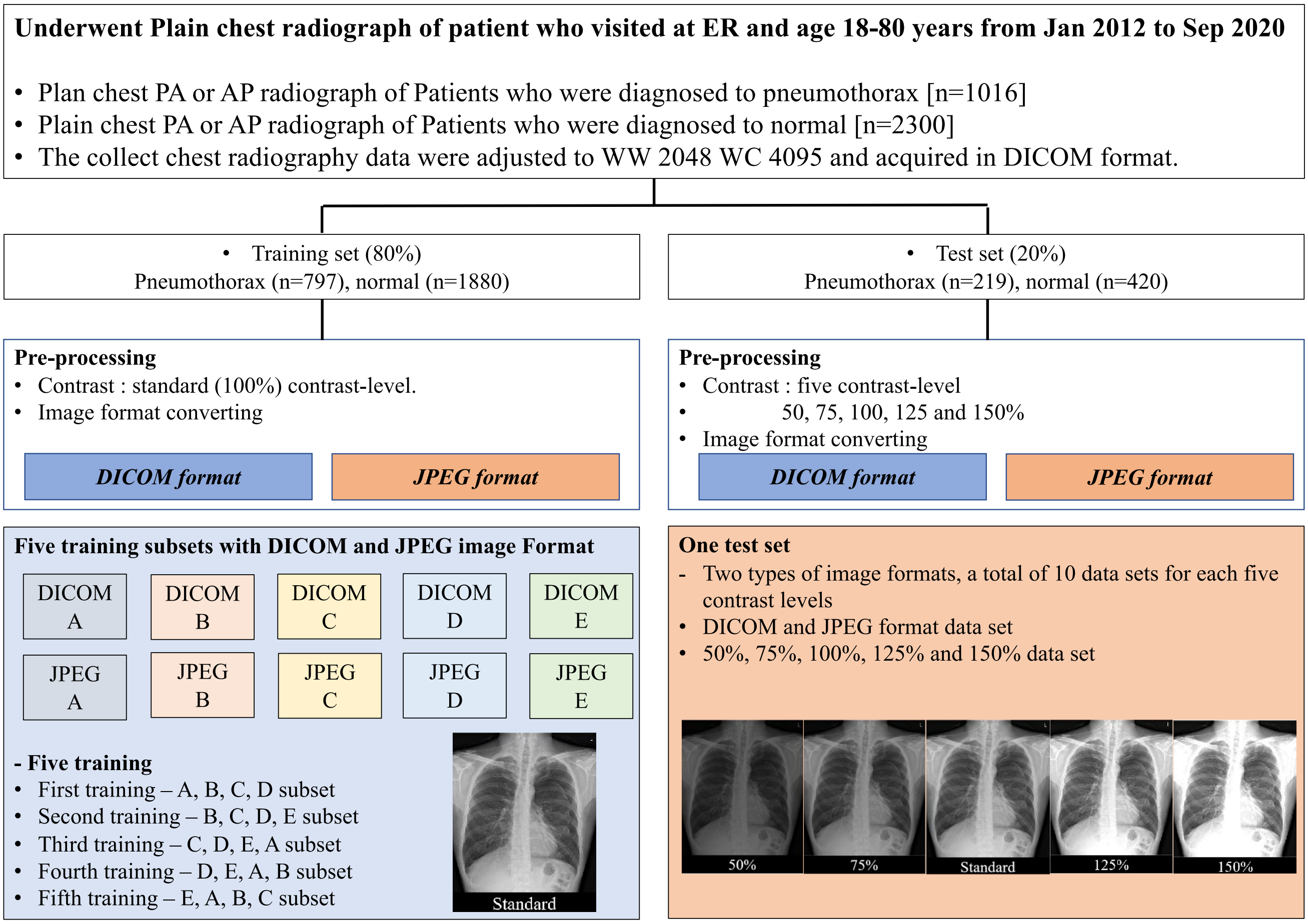

Fig. 1 Flow chart of data collection and analysis in research for pneumothorax detection with deep learning algorithms

Archiving and Communication System (PACS, PiView, Infinitt Healthcare, Seoul, Korea) using a custom-built automated image retrieval system. When storing photographs for data collection, no personal information was included, and no personal identifiable information was collected. In addition, the images were allocated arbitrary numbers and then coded and maintained. Two emergency medicine physicians reached consensus and categorized the pneumothorax size into three groups as follows: (1) small, the separation between the visceral and parietal pleura was less than approximately 1 cm and was typically confined to one lobe of the lung; (2) moderate, the separation was approximately 1–2 cm, often involving more than one lobe or area of the lung; and (3) large, the separation was greater than 2 cm and involved multiple areas of the affected lung [20].

Data Preprocessing and Experiment

All stored images in DICOM format were set to 2048 of window center (WC) and 4095 of window width (WW), that is the standard contrast level (100%) of chest radiograph of the hospital that provided data. Data were randomly separated into the 8:2 training:test set and the training set was divided into five subsets (A, B, C, D, and E sets). The contrast of images in the test set was changed to five levels (50%, 75%, 100%, 125%, and 150% of standard contrast). The WC/WW of each contrast level was set as follows: 1024/2048 (50%), 1536/3071 (75%), 2048/4095 (100%), 2560/5118 (125%), and 3071/6142 (150%). All images were stored as 16-bit DICOM grayscale and 8-bit Joint Photographic Experts Group (JPEG) grayscale images after down sampling using MATLAB (2019a, MathWorks. USA) program. We trained the model using images in each DICOM and JPEG format with a standard contrast level in four (A to D) of five sets (A to E) of the training data. We then tested the trained model with images according to the five contrast levels in the test set. Next, we performed this process with B to E of the training set and test set and repeated it five times for cross-validation.

Network of the Deep Learning Model for the Detection of Pneumothorax

In this study, we applied ResNet-50, which is a widespread deep learning architecture selected by 55% of the top 50 teams in the 2019 US Pneumothorax Detection and Segmentation Competition [21]. They used the ResNet architecture to develop a deep learning-based pneumothorax detection algorithm for chest radiographs. Due to the difficulty of gradients propagation, simply deepening the layers of the neural network did not improve the performance of the model. ResNet aims to solve the optimization issue of deep networks by learning residuals through skip connections. Instead of learning the function directly, the residual block easily alleviates the gradient problem as it learns only the residuals. Even deeper networks can be effectively trained [31]. The overall workflow of our method is illustrated in Fig. 2. We configured 224 × 224 pixels for image size, 100 epochs, and 32 batches. In addition, a stochastic gradient descent optimizer was used, and the learning rate was set to 0.0001 with a weight decay factor of 0.001. The proposed method was implemented using the PyTorch framework. The specifications were a GeForce GTX TITAN RTX 24 GB GPU, i7-4790 (3.60 GHz) central processing unit, and 32 GB RAM for our implementation.

Primary Outcomes

To evaluate and compare the overall performance of the model with that of the proposed method, we used standard metrics, including accuracy, sensitivity, specificity, and F1-score. The following metrics are based on the predicted results and their ground truth as true positive (TP), false positive (FP), true negative (TN), and false negative (FN). In our experiments, TP refers to a chest radiograph with a diagnosed pneumothorax as "pneumothorax." TN, in contrast, refers to a correctly diagnosed chest radiograph without pneumothorax as "normal." FN refers to chest radiograph with an existing pneumothorax as a misdiagnosed "normal," while FP presents the result the other way around. We used the receiver-operating characteristic (ROC) curve for the evaluation of our model. Because chest radiograph is widely used and available screening tool for the detection of pneumothorax in emergency and admission room than a confirmation tool, we set the optimal cut-off values that had the highest sensitivity and a specificity of 0.5 or greater when testing the model with standard contrast level (100%) images. We calculated the area under the ROC curve (AUC), sensitivity, specificity, accuracy, and F1-score with these cut-off values in each test with images of 50%, 75%, 125%, and 150% contrast levels using the following equations:

The AUC denotes the area under the ROC curve that plots the relationship between the TP rate (sensitivity) and the false positive rate (1-specificity).

Sensitivity (TP rate, recall) refers to the likelihood of the positive samples if the condition is actually predicted:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

Specificity (TN rate) refers to the likelihood of the negative samples as the following condition:

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

Accuracy is used as a measure of the proportion of correctly predicted samples out of the total number of predictions as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

The F1-score is the harmonic mean of precision (the probability of true positives among true and false positives), and recall is as follows:

$$F1\ score = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

Statistical Analysis

We performed data organization using Excel (Microsoft, Redmond, WA, USA) and analyzed the organized data using NCSS (2020, LLC. Kaysville, UT, USA). Using Kolmogorov–Smirnov tests, the normal distribution of all datasets was used. For categorical data, the statistics in this paper are presented as frequency and percentage. The pneumothorax and normal data groups were compared using the independent *t*-test or the Mann–Whitney *U* test for continuous variables, and the chi-squared test was used for categorical variables. We used a single ROC curve and cut-off analysis for each test and two ROC curves with the paired group design for comparing

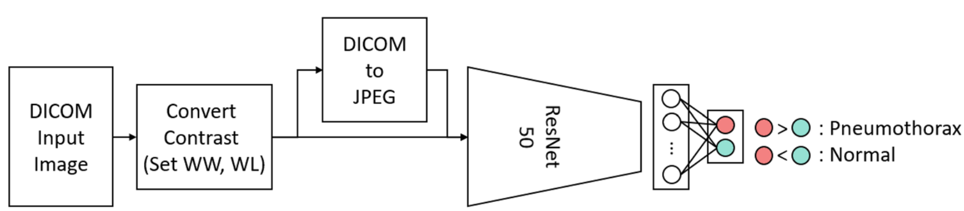

Fig. 2 Overview of detailed workflow used to extract the features of the image. In our workflow, ResNet-50 [31] was utilized as our backbone networks

the ROC curves between each contrast levels in each format, between DICOM and JPEG formats at same contrast level. Statistical significance was set at $p < 0.05$.

Results

Experimental Setup

A total of 3316 images, containing 1016 pneumothorax images (386 small size, 281 moderate size, and 349 large size cases) and 2300 normal images, were collected from 797 and 1812 patients, respectively; the characteristics of the data are shown in Table 1.

Overall Performance for Pneumothorax Detection with Chest Radiography According to Five Contrast Levels and Image Formats

The ROC curves for the overall performance of this network through five tests according to the contrast level and format of the images are shown in Fig. 3. AUCs (95% CI) of the test with standard contrast (100%) images were 0.999 (0.998–0.999) for DICOM format and 0.998 (0.995–0.999) for JPEG format after training with the same-level images in the training set. The value of the test with 50%, 75%, 125%, and 150% of contrast level images and DICOM format were 0.980 (0.976–0.984), 0.987 (0.984–0.990), 0.959 (0.950–0.965), and 0.929 (0.917–0.938), respectively. The value of the test with 50%, 75%, 125%, and 150% of contrast level images and JPEG format were 0.976 (0.971–0.980), 0.994 (0.991–0.996), 0.964 (0.957–0.971), and 0.960 (0.953–0.967), respectively. In comparison between each contrast level in both DICOM and JPEG formats, values of AUC were significantly different (all $p < 0.001$), except the comparison between 125 and 150% in JPEG format ($p = 0.382$). In comparison between DICOM and JPEG formats at each contrast level, values of AUC were significantly different (all $p < 0.001$) except the comparison between 50% ($p = 0.079$) and 100% ($p = 0.082$).

Table 1 Baseline characteristics of participants who provided images and data matrix according to pneumothorax size in datasets

Dataset ($n = 3316$)	Pneumothorax ($n = 1016$)			Normal ($n = 2300$)	p-value
Age, mean [SD]	42.8 [19.4]			43.1 [10.4]	0.478
Sex, male, n (%)	786 (77.4)			1,390 (60.4)	< 0.001*
Pneumothorax size	Small ($n = 386$)	Moderate ($n = 281$)	Large ($n = 349$)		
Train set (80%)	308	212	277	1880	
Test set (20%)	78	69	72	420	

SD standard deviation; continuous data are reported as mean (standard deviation) while categorical variables are presented as n (percentage). The independent t-test was used to compare pneumothorax and normal data groups. The chi-square test was used to examine categorical variables that are provided as numbers and percentages. Statistical significance was set at $p < 0.05$. Pneumothorax size were as follows: (1) small, the separation between the visceral and parietal pleura was less than approximately 1 cm and was typically confined to one lobe of the lung; (2) moderate, separation was approximately 1–2 cm, and often involved more than one lobe or area of the lung; and (3) large, separation was greater than 2 cm and involved multiple areas of the affected lung

Comparison of Average Outcomes According to 50%, 75%, 100%, 125%, and 150% of Contrast Levels in Each DICOM and JPEG Formats Over Five Tests

Optimal cut-off values, which had the highest sensitivity with ≥ 0.5 of specificity, were 0.238 [0.047] for DICOM and 0.218 [0.028] for JPEG in each test with images of 100% contrast level.

Confusion matrices and outcomes for the diagnostic performance of each test are shown in Fig. 4 and Table 2, respectively.

Comparison of Performance According to 5-Contrast Levels in DICOM Format

Values of AUC were decreased from 0.999 [0.004] of 100% to 0.992 [0.004] of 75%, 0.979 [0.010] of 50%, 0.964 [0.020] of 125%, and 0.918 [0.079] of 150% by contrast level reduction and gain. The sensitivity values were maintained more than about 0.970, whereas the specificity decreased from 0.998 [0.001] for 100% contrast level to 0.269 [0.247] for 50% contrast level and 0.372 [0.378] for 150% contrast level. The accuracy and F1-scores also decreased with these changes.

Comparison of Performance According to 5-Contrast Levels in JPEG Format

Values of AUC were decreased from 0.998 [0.001] for 100% to 0.996 [0.001] for 75%, 0.978 [0.008] for 50%, 0.969 [0.020] for 125%, and 0.960 [0.022] for 150% by contrast level reduction and gain. Values of sensitivity were maintained well over approximately 0.970, whereas specificity was decreased from 0.998 [0.001] at 100% contrast level to 0.153 [0.182] at 50% contrast level and 0.542 [0.166]

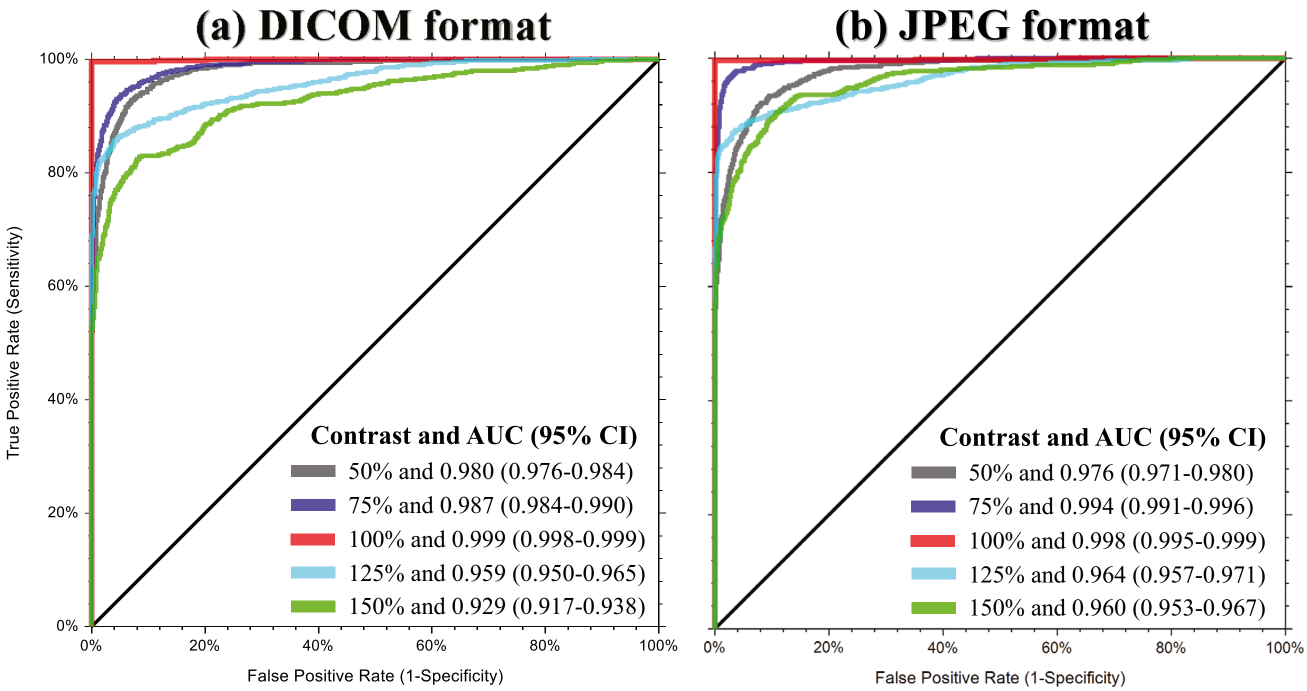

Fig. 3 Receiver-operator characteristics (ROC) curves of overall performance for the detection of pneumothorax using ResNet-50 network with chest radiograph over five tests according to 50%, 75%, 100%, 125%, and 150% contrast level and image formats. (**a**) DICOM format, (**b**) JPEG format

at 150% contrast level. The accuracy and F1-scores also decreased with these changes.

Examples of prediction success and failure based on contrast changes are shown in Fig. 5. Reduction or gain of contrast level influenced the prediction for pneumothorax with normal images more than the prediction for normal with pneumothorax images in both image formats.

Discussion

We trained a deep learning model for the pneumothorax detection with chest radiographs using a ResNet-50 network and investigated the performance of the model in the test set according to contrast levels and formats of images after training. The performance of the deep learning algorithm could be decreased when tested with images of different contrast levels from images in training, and it might differ according to the image formats.

Recently, several studies have demonstrated high performance in pneumothorax detection using deep learning. Park et al. reported that the performance was 0.985 of AUC among five multiclass lesions using 15,809 images from two hospitals [17] and Hwang et al. demonstrated that the performance was 0.965–0.979 of AUC with external validation data from five institutions after training the four major thoracic diseases and normal with approximately 100,000 images from one hospital [18]. Majkowska et al. reported that the performance of pneumothorax detection among four classifications was 0.940 ~ 0.950 of AUC in a test with National Institutes of Health (NIH) ChestX-ray14 after training with 1818 images from one center [15]. Wang et al. improved the performance to 0.991 of the AUC using real-time augmentation to solve imbalanced data [14]. However, they did not mention the size of pneumothorax in the dataset and could not determine the size effect on the performance of pneumothorax detection. Taylor et al. reported the performance according to pneumothorax size and showed that the AUC was 0.940 for moderate-to-large amounts of pneumothorax in external validation with NIH ChestX-ray14 using the ResNet network [20]. Yi et al. showed that sensitivity of the performance was inversely correlated with pneumothorax size [16]. In this study, using the ResNet-50 network, AUCs were 0.999 [0.004] in DICOM format and 0.998 [0.001] in JPEG format for the detection of pneumothorax among two classifications in a test with standard contrast level images after training with same-level images obtained from one hospital. Although our data included 386 small images with moderate and large numbers, the performance was high. However, we did not know the maintenance of performance because we did not test the external validation with other datasets.

Chest radiography is the most common modality used for observing the long-scale contrast area because it uses a high

(a) DICOM

		Actual P	Actual N
Predicted	P	1093	866
	N	2	1234

50 % contrast level

		Actual P	Actual N
Predicted	P	1095	793
	N	0	1307

75 % contrast level

		Actual P	Actual N
Predicted	P	1089	4
	N	6	2096

100 % contrast level

125 % contrast level

		Actual P	Actual N
Predicted	P	1083	929
	N	12	1171

150 % contrast level

		Actual P	Actual N
Predicted	P	1010	673
	N	85	1427

(b) JPEG

		Actual P	Actual N
Predicted	P	1093	806
	N	2	1294

50 % contrast level

		Actual P	Actual N
Predicted	P	1095	803
	N	0	1297

75 % contrast level

		Actual P	Actual N
Predicted	P	1091	19
	N	4	2081

100 % contrast level

125 % contrast level

		Actual P	Actual N
Predicted	P	1078	937
	N	17	1163

150 % contrast level

		Actual P	Actual N
Predicted	P	1065	575
	N	30	1525

Fig. 4 Confusion matrices of the test performance for the detection of pneumothorax using ResNet-50 network with five contrast level images (50%, 75%, 100%, 125%, and 150%) in (**a**) DICOM and (**b**) JPEG. P, positive; N, negative

Table 2 Performance outcomes according to five contrast levels in each DICOM and JPEG formats over five tests for the detection of pneumothorax using ResNet-50 network with chest radiography

	DICOM					JPEG format				
	50%	75%	100%	125%	150%	50%	75%	100%	125%	150%
Cut-off, mean [SD]	0.238 [0.047]					0.218 [0.028]				
AUC, mean [SD]	0.979 [0.010]	0.992 [0.004]	0.999 [0.004]	0.964 [0.020]	0.918 [0.079]	0.978 [0.008]	0.996 [0.001]	0.998 [0.001]	0.969 [0.020]	0.960 [0.022]
Sensitivity, mean [SD]	0.999 [0.002]	1.000 [0.004]	0.995 [0.002]	0.968 [0.068]	0.991 [0.004]	1.000 [0.000]	1.000 [0.002]	0.995 [0.000]	0.973 [0.004]	0.982 [0.017]
Specificity, mean [SD]	0.269 [0.247]	0.579 [0.117]	0.998 [0.001]	0.813 [0.133]	0.372 [0.378]	0.153 [0.182]	0.590 [0.137]	0.998 [0.001]	0.751 [0.157]	0.542 [0.166]
Accuracy, mean [SD]	0.520 [0.162]	0.723 [0.076]	0.997 [0.000]	0.846 [0.075]	0.585 [0.248]	0.444 [0.119]	0.730 [0.089]	0.997 [0.000]	0.816 [0.099]	0.693 [0.114]
F1-score, mean [SD]	0.597 [0.086]	0.790 [0.064]	0.995 [0.001]	0.790 [0.064]	0.655 [0.195]	0.553 [0.059]	0.690 [0.022]	0.995 [0.000]	0.788 [0.100]	0.704 [0.079]

SD standard deviation

voltage of 90 to 130 kVp [32–34] and can be used to diagnose pneumothorax with a small contrast difference between tissues. However, there is no definite standard for proper contrast in chest radiographs for pneumothorax detection when physicians interpret images, and they usually control the contrast level to detect pneumothorax well. In DICOM images, the WW and WC play a role in adjusting the contrast level of various human tissues. In addition, each radiograph image has a different contrast because even if the same radiograph is irradiated, the characteristics and brightness of the image are determined by various variables, such as radiation dose, scattering radiation, patient thickness, equipment, and image post-processing [35–40]. Owing to the two problems mentioned above, it is impossible to maintain the same contrast according to the characteristics of hospitals and patients on chest radiographs. As a result of our study, the performance was high in the test with the same contrast level images after training with images of the standard level; however, when tested with different contrast images with the training data, the performance deteriorated depending on the contrast. Some studies with excellent performance for pneumothorax detection might have used images of similar contrast levels during training and testing. Wang et al. reported that histogram equalization could improve the performance of a model in a typical image format [14]. The difference in performance according to the contrast level can be solved to some extent by normalization or histogram equalization.

Chest radiographs are stored in DICOM, a standard medical imaging method, and transmitted through PACS [41, 42]. DICOM contains sensitive personal information about the patients [43, 44]. In addition, 12- to 16-bit DICOM formats contain more information than 8-bit JPEG formats [45, 46]. Down sampling the image to JPEG reduces the data size to 1/10 but discards a lot of information [46]. The original pixel values were lost during the image compression process in the JPEG format. Maruyama et al. reported that low-quality general image formats affect machine learning algorithms [47]. Their study mentioned that the JPEG image file format reduces the classification accuracy owing to data loss. When analyzing previous studies related to deep learning in pneumothorax, it was found that there was no information on the image format, or that the data were collected using DICOM and down sampling was performed with JPEG or PNG. Information for contrast level and image format is described in Table 3 [14–20]. Kim et al. reported that the performance of classification of cardiomegaly on chest radiographs was no different between 16-bit Tagged Image File Format (TIFF) and 8-bit PNG/JPEG [48]. In our study, the performance also was no different between DICOM and JPEG formats in the test with images of 100% contrast level. However, this could be different when the contrast level of the images in the test set was changed.

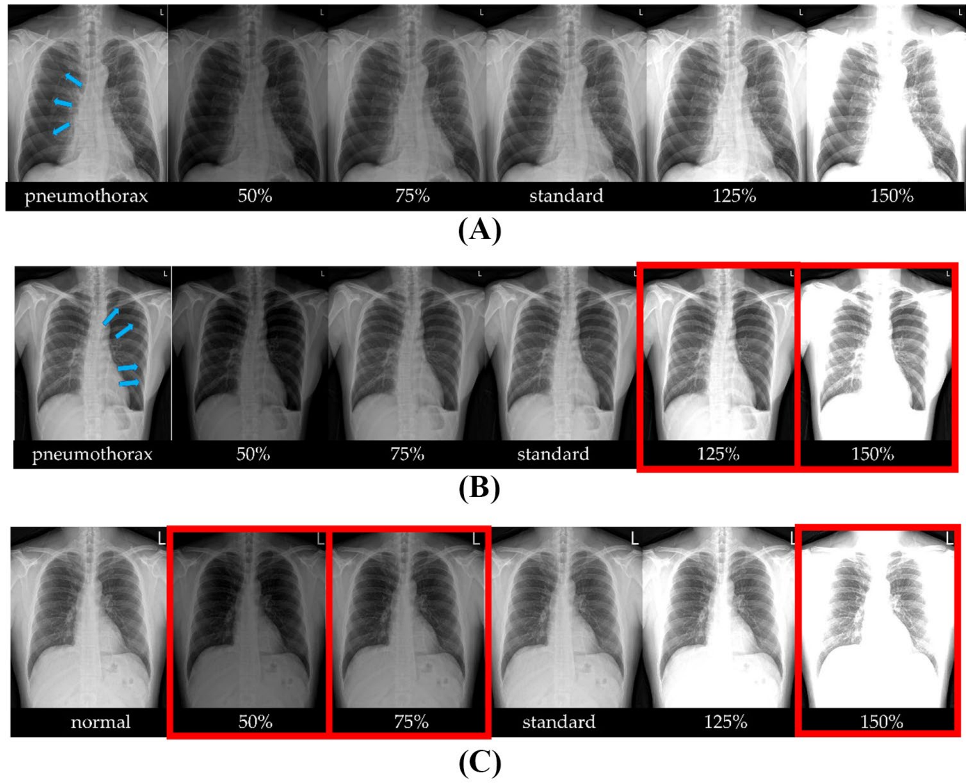

Fig. 5 Examples of right and wrong detection of pneumothorax using ResNet-50 network according to contrast level of the chest radiograph. **A** On chest radiography with a pneumothorax in the right lung, all images were detected to pneumothorax according to all contrast levels. **B** On chest radiography with a pneumothorax in the left lung, standard and dark images of 50% and 75% contrast levels were detected as pneumothorax, but light images of 125% and 150% contrast levels were predicted as normal images. **C** On chest radiography with normal images, images of 50%, 75%, and 150% contrast levels were predicted as pneumothorax. Blue arrows indicate the visceral pleural line of the pneumothorax and red boxes indicate images of false positive or negative

Table 3 Summary of previous studies on the detection of pneumothorax using deep learning

Author	Year	Architecture	Diseases of dataset	Classification	Pneumothorax size	AUC	Image format
Wang et al. [14]	2021	ChestNet	Atelectasis, emphysema, pulmonary congestion, pneumothorax, enlargement of heart, effusion, pneumonia, and infiltration Normal	Pneumothorax Others	-	0.991	PNG
Majkowska et al. [15]	2020	Xception	Pneumothorax, opacity, nodule or mass, fracture Others	Pneumothorax Others	-	0.940 ~ 0.950	PNG
Yi H. et al. [16]	2020	ResNet 152	Pneumothorax	Pneumothorax Without pneumothorax	Small Moderate Large	0.840	PNG
Park et al. [17]	2020	Unknown	Normal Interstitial opacity, pleural effusion, pneumothorax	Normal Abnormal	-	0.985	Unknown
Hwang et al. [18]	2019	DenseNet	Normal Malignancy, tuberculosis, pneumonia, pneumothorax	Normal Any target disease	-	0.965–0.979	DICOM
Park et al. [19]	2019	YOLO	Pneumothorax	Pneumothorax Without pneumothorax	-	0.984	PNG
Taylor et al. [20]	2018	ResNet VGG 19 Inception Xception	Pneumothorax	Pneumothorax Without pneumothorax	Small Moderate Large	0.950 0.970 0.970 0.980	JPEG

In future work, we will investigate the effect of the contrast level and image formats by various neural network in different diseases and modality. At the same time, we will carry out in-depth research around proper augmentation such as histogram equalization and normalization with optimal window level and width applying it to the current medical image and deep learning system. Finally, we will validate these results with external datasets.

Limitation

The performance changes due to format and image contrast level factor were compared only in the pneumothorax detection task. The effects of these factors on various diseases were not tested. Second, we performed pneumothorax classification on single institution dataset. This paper does not include experiments on external data. Third, we did not evaluate the performance change of the model in more diverse file formats such as PNG or TIFF. The compression algorithms for each format are different, but we did not evaluate how these compression algorithms affect the deep learning model. Finally, we applied ResNet-50 network model, which was the best performance of detection for pneumothorax; however, we did not investigate the change of performance according to contrast level and image format when other networks would be used.

Conclusions

The contrast level and format of medical images could influence the performance of the deep learning model. It is required to train with various contrast levels and formats of image, and further image processing techniques for improvement and maintenance of the proposed model performance.

Abbreviations DICOM: Digital Imaging and Communication in Medicine; JPEG: Joint Photographic Experts Group; CNN: Convolution neural network; ROC curve: Receiver-operating characteristic curve; PACS: Picture Archiving and Communication System; SGD: Stochastic gradient descent; CAD: Computer-aided diagnosis; GPU: Graphics processing unit; SNR: Signal-to-noise ratio; IRB: Institutional Review Board; TP: True positive; TN: True negative; FP: False positive; FN: False negative

Author Contribution Conceptualization, M.S.Y, G.K, D.H., and J.O.; data curation, M.S.Y. and G.K.; formal analysis, B.K., D.H., J.L., and J.L.; funding acquisition, J.O.; methodology, G.K, J.L., and J.R.; supervision, J.R. and J.O.; validation, M.S.Y. and G.K.; visualization, G.K. and J.R.; writing—original draft, M.S.Y. and J.R.; writing—review and editing, M.S.Y., J.R., G.K., and J.O. All authors have read and agreed to the published version of the manuscript.

Funding This study was supported by the National Research Foundation of Korea (NRF-2022R1A2C1012627).

Data Availability The data presented in this study are available on request from the corresponding author.

Declarations

Ethics Approval This study was approved by the Institutional Review Board (IRB) of Hanyang University Hospital (ref. no. HYUH 2020–03-039).

Consent to Participate The requirement for informed consent was waived by the IRBs of Hanyang University Hospital. All methods and procedures were carried out in accordance with the Declaration of Helsinki.

Consent for Publication This manuscript has not been published or presented elsewhere in part or in entirety and is not under consideration by another journal. We have read and understood your journal's policies, and we believe that neither the manuscript nor the study violates any of these.

Conflict of Interest The authors declare no competing interests.

References

1. Mintz Y, Brodie R: Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 28(2):73-81, 2019
2. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL: Artificial intelligence in radiology. Nat Rev Cancer 18(8):500-510, 2018
3. Syed AB, Zoga AC: Artificial Intelligence in Radiology: Current Technology and Future Directions. Semin Musculoskelet Radiol 22(5):540-545, 2018
4. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP: Artificial intelligence applications for thoracic imaging. Eur J Radiol 123:108774, 2020
5. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, et al: Deep Learning-Based Automatic Detection Algorithm Development and Evaluation Group. Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. Clin Infect Dis 69(5):739–747, 2019
6. Hwang EJ, Nam JG, Lim WH, Park SJ, Jeong YS, Kang JH, et al: Deep Learning for Chest Radiograph Diagnosis in the Emergency Department. Radiology 293(3):573-580, 2019
7. Lee SM, Seo JB, Yun J, Cho YH, Vogel-Claussen J, Schiebler ML, et al: Deep Learning Applications in Chest Radiography and Computed Tomography: Current State of the Art. J Thorac Imaging 34(2):75-85, 2019
8. Arshad H, Young M, Adurty R, Singh AC: Acute Pneumothorax. Crit Care Nurs Q 39(2):176-89, 2016
9. Thomsen L, Natho O, Feigen U, Schulz U, Kivelitz D: Value of digital radiography in expiration in detection of pneumothorax. Rofo 186(3):267-73, 2014
10. Seow A, Kazerooni EA, Pernicano PG, Neary M: Comparison of upright inspiratory and expiratory chest radiographs for detecting pneumothoraces. AJR Am J Roentgenol 166(2):313-6, 1996
11. Carr JJ, Reed JC, Choplin RH, Pope TL Jr, Case LD: Plain and computed radiography for detecting experimentally induced pneumothorax in cadavers: implications for detection in patients. Radiology 183(1):193-9, 1992

12. Chen KY, Jerng JS, Liao WY, Ding LW, Kuo LC, Wang JY, et al: Pneumothorax in the ICU: patient outcomes and prognostic factors. Chest 122(2):678-83., 2002
13. Hallifax R, Janssen JP: Pneumothorax-Time for New Guidelines? Semin Respir Crit Care Med 40(3):314-322, 2019
14. Wang Y, Sun L, Jin Q. Enhanced Diagnosis of Pneumothorax with an Improved Real-Time Augmentation for Imbalanced Chest X-rays Data Based on DCNN. IEEE/ACM Trans Comput Biol Bioinform 18(3):951-962, 2021
15. Majkowska A, Mittal S, Steiner DF, Reicher JJ, McKinney SM, Duggan GE, et al: Chest Radiograph Interpretation with Deep Learning Models: Assessment with Radiologist-adjudicated Reference Standards and Population-adjusted Evaluation. Radiology 294(2):421-431, 2020
16. Yi PH, Kim TK, Yu AC, Bennett B, Eng J, Lin CT: Can AI outperform a junior resident? Comparison of deep neural network to first-year radiology residents for identification of pneumothorax. Emerg Radiol 27(4):367-375, 2020
17. Park S, Lee SM, Lee KH, Jung KH, Bae W, Choe J, et al: Deep learning-based detection system for multiclass lesions on chest radiographs: comparison with observer readings. Eur Radiol 30(3):1359-1368, 2020
18. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, et al: DLAD Development and Evaluation Group. Development and Validation of a Deep Learning-Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. JAMA Netw Open 2(3):e191095, 2019
19. Park S, Lee SM, Kim N, Choe J, Cho Y, Do KH, et al: Application of deep learning-based computer-aided detection system: detecting pneumothorax on chest radiograph after biopsy. Eur Radiol 29(10):5341-5348, 2019
20. Taylor AG, Mielke C, Mongan J: Automated detection of moderate and large pneumothorax on frontal chest Radiographs using deep convolutional neural networks: Aretrospective study. PLoS Med 15(11):e1002697, 2019
21. Tolkachev A, Sirazitdinov I, Kholiavchenko M, Mustafaev T, Ibragimov B: Deep Learning for Diagnosis and Segmentation of Pneumothorax: The Results on the Kaggle Competition and Validation Against Radiologists. IEEE J Biomed Health Inform 25(5):1660-1672, 2021
22. Bae J, Yu S, Oh J, Kim TH, Chung JH, Byun H, et al: External Validation of Deep Learning Algorithm for Detecting and Visualizing Femoral Neck Fracture Including Displaced and Non-displaced Fracture on Plain Radiograph. J Digit Imaging 34(5):1099-1109, 2021
23. Geirhos R, Janssen DHJ, Schütt HH, Rauber J, Bethge M, Wichmann FA. Comparing deep neural networks against humans: object recognition when the signal gets weaker. ArXiv170606969 Cs Q-Bio Stat [Internet]. 2017 Jun 21 [cited 2017 Dec 31]; Available from: http://arxiv.org/abs/1706.06969
24. Akbarinia A, Gil-Rodríguez R: Deciphering image contrast in object classification deep networks. Vision Res 173:61-76, 2020
25. Geirhos, R, Temme, CRM., Rauber, J, Schütt HH., Bethge M, Wichmann FA: Generalisation in humans and deep neural networks. Adv. Neural Inf. Process. Syst., 31: 7549-61, 2018
26. Giełczyk A, Marciniak A, Tarczewska M, Lutowski Z. Preprocessing methods in chest X-ray image classification. PLoS One. 17(4):e0265949, 2022
27. Li MD, Arun NT, Aggarwal M, Gupta S, Singh P, Little BP, et al. Improvement and Multi-Population Generalizability of a Deep Learning-Based Chest Radiograph Severity Score for COVID-19. 101(29): e29587, 2022
28. Thian YL, Ng DW, Hallinan JTPD, Jagmohan P, Sia SY, Mohamed JSA, Quek ST et al. Effect of Training Data Volume on Performance of Convolutional Neural Network Pneumothorax Classifiers. J Digit Imaging. 35(4):881-892, 2022
29. Wang F, Casalino LP, Khullar D: Deep Learning in Medicine-Promise, Progress, and Challenges. JAMA Intern Med 179(3):293-294, 2019
30. Sarvamangala DR, Kulkarni RV: Convolutional neural networks in medical image understanding: a survey. Evol Intell :1–22, 2021.
31. He F, Liu T, Tao D. Why ResNet Works? Residuals Generalize. IEEE Trans Neural Netw Learn Syst. 31(12):5349-5362, 2020
32. Sauter AP, Andrejewski J, De Marco F, Willer K, Gromann LB, Noichl W, et al: Optimization of tube voltage in X-ray dark-field chest radiography. Sci Rep. 9(1):8699, 2019
33. Nocetti D, Ubeda C, Calcagno S, Acevedo J, Pardo D: Comparison of image quality among three Radiograph systems for chest radiography: first step in optimisation. Radiat Prot Dosimetry 165(1-4):386-91, 2015
34. Grewal RK, Young N, Colins L, Karunnaratne N, Sabharwal N: Digital chest radiography image quality assessment with dose reduction. Australas Phys Eng Sci Med 35(1):71-80, 2012
35. Muhogora W, Padovani R, Msaki P: Initial quality performance results using a phantom to simulate chest computed radiography. J Med Phys 36(1):22-8, 2011
36. Zheng X: Patient size based guiding equations for automatic mAs and kVp selections in general medical radiograph projection radiography. Radiat Prot Dosimetry 174(4):545-550, 2017
37. Veldkamp WJ, Kroft LJ, Geleijns J: Dose and perceived image quality in chest radiography. Eur J Radiol 72(2):209-17, 2009
38. Tischenko O, Hoeschen C, Buhr E: Reduction of anatomical noise in medical Radiograph images. Radiat Prot Dosimetry 114(1-3):69-74, 2005
39. Mouzai M, Tarabet C, Mustapha A: Low-contrast Radiograph enhancement using a fuzzy gamma reasoning model. Med Biol Eng Comput 58(6):1177-1197, 2020
40. Schaefer-Prokop C, Neitzel U, Venema HW, Uffmann M, Prokop M: Digital chest radiography: an update on modern technology, dose containment and control of image quality. Eur Radiol 18(9):1818-30, 2008
41. Peer S, Vogl R, Peer R, Jaschke W: Sophisticated hospital information system/radiology information system/picture archiving and communications system (PACS) integration in a large-scale traumatology PACS. J Digit Imaging 12(2 Suppl 1):99-102, 1999
42. Law MY, Zhou Z: New direction in PACS education and training. Comput Med Imaging Graph 27(2-3):147-56, 2003
43. Varma DR: Managing DICOM images: Tips and tricks for the radiologist. Indian J Radiol Imaging 22(1):4-13, 2012
44. Larobina M, Murino L. Medical image file formats. J Digit Imaging. 27(2):200-6, 2014
45. Wang B, Li YB, Zhao JX, Sui X, Kong XW: JPEG compression history detection based on detail deviation. Math Biosci Eng 16(5): 5584-5594, 2019
46. Graham RN, Perriss RW, Scarsbrook AF: DICOM demystified: a review of digital file formats and their use in radiological practice. Clin Radiol 60(11):1133-1140, 2005
47. Maruyama T, Hayashi N, Sato Y, Hyuga S, Wakayama Y, Watanabe H, et al: Comparison of medical image classification accuracy among three machine learning methods. J Xray Sci Technol 26(6):885-893, 2018
48. Kim MJ, Kim JH: development of convolutional neural network model for classification of cardiomegaly X-ray images. Journal of Mechanics in Medicine and Biology 22(8):1-14, 2022