

## Article

# Development of Chest X-ray Image Evaluation Software Using the Deep Learning Techniques

Kousuke Usui <sup>1</sup>, Takaaki Yoshimura <sup>2,3,4,5</sup> , Shota Ichikawa <sup>6,7</sup>  and Hiroyuki Sugimori <sup>4,5,8,\*</sup> 

- <sup>1</sup> Graduate School of Health Sciences, Hokkaido University, Sapporo 0600812, Japan  
<sup>2</sup> Department of Health Sciences and Technology, Faculty of Health Sciences, Hokkaido University, Sapporo 0600812, Japan  
<sup>3</sup> Department of Medical Physics, Hokkaido University Hospital, Sapporo 0608648, Japan  
<sup>4</sup> Global Center for Biomedical Science and Engineering, Faculty of Medicine, Hokkaido University, Sapporo 0608648, Japan  
<sup>5</sup> Clinical AI Human Resources Development Program, Faculty of Medicine, Hokkaido University, Sapporo 0608648, Japan  
<sup>6</sup> Department of Radiological Technology, School of Health Sciences, Faculty of Medicine, Niigata University, Niigata 9518518, Japan  
<sup>7</sup> Institute for Research Administration, Niigata University, Niigata 9502181, Japan  
<sup>8</sup> Department of Biomedical Science and Engineering, Faculty of Health Sciences, Hokkaido University, Sapporo 0600812, Japan  
\* Correspondence: sugimori@hs.hokudai.ac.jp; Tel.: +81-11-706-3410

**Abstract:** Although the widespread use of digital imaging has enabled real-time image display, images in chest X-ray examinations can be confirmed by the radiologist's eyes. Considering the development of deep learning (DL) technology, its application will make it possible to immediately determine the need for a retake, which is expected to further improve examination throughput. In this study, we developed software for evaluating chest X-ray images to determine whether a repeat radiographic examination is necessary, based on the combined application of DL technologies, and evaluated its accuracy. The target population was 4809 chest images from a public database. Three classification models (CLMs) for lung field defects, obstacle shadows, and the location of obstacle shadows and a semantic segmentation model (SSM) for the lung field regions were developed using a fivefold cross validation. The CLM was evaluated using the overall accuracy in the confusion matrix, the SSM was evaluated using the mean intersection over union (mIoU), and the DL technology-combined software was evaluated using the total response time on this software (RT) per image for each model. The results of each CLM with respect to lung field defects, obstacle shadows, and obstacle shadow location were 89.8%, 91.7%, and 91.2%, respectively. The mIoU of the SSM was 0.920, and the software RT was  $3.64 \times 10^{-2}$  s. These results indicate that the software can immediately and accurately determine whether a chest image needs to be re-scanned.

**Keywords:** deep learning; chest X-ray image; quality assurance system



**Citation:** Usui, K.; Yoshimura, T.; Ichikawa, S.; Sugimori, H. Development of Chest X-ray Image Evaluation Software Using the Deep Learning Techniques. *Appl. Sci.* **2023**, *13*, 6695. <https://doi.org/10.3390/app13116695>

Academic Editors: Zhonghua Sun and Adam Konefał

Received: 21 April 2023

Revised: 26 May 2023

Accepted: 30 May 2023

Published: 31 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Several X-ray examinations are performed annually on a global scale [1], of which chest X-ray (CXR) examination accounts for the largest number of cases [2]. In the past, the X-ray intensity distribution obtained by irradiating the subject with X-rays was taken on film, and the image could be displayed by developing the distribution (screen/film system). However, in recent years, digital imaging methods, such as computed radiography (CR), have been increasingly used. As a result, the time required to display the taken images has been significantly shortened. Although imaging plates require time for reading and processing, flat panel detectors (FPDs), which display images in real time, are becoming more widely used.

In these X-ray examinations, the target area is required to be accurately visualized, and when it cannot be visualized, retaking is necessary, which contributes to an increase in examination time and radiation exposure experienced by the subject. The study by Lin et al. [3] reported the factors that cause retakes in general X-ray examinations, including CXR examination, and most of them were positioning errors, defects of necessary parts, and artifacts. Although the widespread use of digital imaging has enabled real-time image display, the radiologist's eyes are still used to determine the need for a retake. Therefore, to further improve the efficiency and accuracy of X-ray examinations, it is essential to have a system that can immediately determine such needs.

In recent years, several studies have performed medical image analyses using the deep learning (DL) techniques [4–6]. Among the DL techniques using convolutional neural network, classification [7], semantic segmentation [8], and object detection [9] are suitable for medical image analyses, and there are many reports on these techniques. In the field of CXR images, the classification, semantic segmentation, and object detection technologies have been used for lesion classification [10], the semantic segmentation of the lung field areas [11,12], and the detection of diseases in lung field areas [13], respectively. In addition, with the recent global coronavirus disease 2019 pandemic, these technologies can be applied to detect the presence of pneumonia and segment the pneumonia in the lung field area [14,15]. The development of these technologies has led to the development of the technology.

Although these technologies have been improving, their use in determining the need for a retake has not yet been fully elucidated. Konica Minolta, Inc., has already commercialized the AeroDR solution to improve the efficiency of medical examinations, and its CS-7 console is equipped with functions to detect lung field defects and body motion in the frontal CXR images [16]. However, using the classification and semantic segmentation techniques in DL technology, the presence of lung field defects or obstacle shadows on the taken image can be classified, and the location of the obstacle shadows, either inside or outside the lung field, can be determined, regardless of the imaging environment, such as a hospital room or an X-ray room. Junhao et al. [17] applied the DL techniques to the construction of a quality assurance (QA) system for CXR images. Although this QA system can detect lung field defects and artifacts, it does not discriminate between medical and nonmedical devices and does not recognize images in a similar way to humans, making it insufficient to determine the need for a retake. Although the application of deep learning technology to medical imaging has made progress, there are few applications in the field of medical image acquisition technology, and furthermore, no applications have been identified in the field of determining whether or not an image should be retaken.

Therefore, in this study, we developed software for evaluating CXR images to determine whether a CXR image needs to be retaken, based on the combined application of DL techniques, and evaluated its accuracy.

## 2. Materials and Methods

### 2.1. Subjects

In this study, we used the publicly available dataset “ChestX-ray8” provided by the National Institutes of Health (NIH) Clinical Center [18]. The software was based on 4809 images from this dataset. The purpose of this software was to provide an artificial intelligence-based judgment on whether a CXR image should be retaken for radiologists in CXR examinations in clinical settings. Therefore, it is necessary to be able to extract the CXR images which should be retaken from many CXR images.

We used 1000 of 4809 images to create a classification model (CLM). Because each CLM was created independently, we did not consider the overlap of data among the models.

In addition, we created a semantic segmentation model (SSM) to improve the visibility of the lung field region. In this case, we used 1000 chest images without considering the overlap of data with other models. Table 1 shows the breakdown of the 4809 chest images, including duplicates.

**Table 1.** Usage of the images used (4809 images in total).

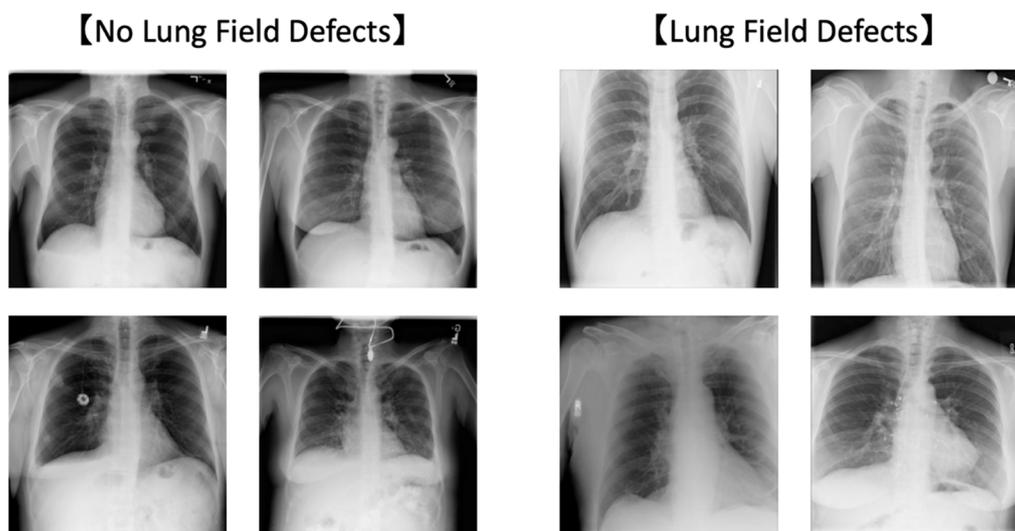
Deep Learning Model	No. of Images
Lung field defect classification model	
OK	1000
NG	1000
Disorder shading classification model	
Medical device	1000
Others (nonmedical device)	1000
None	1000
In-out classification model	
In	(270)
Out	(730)
Semantic segmentation model	1000
Total amount	4809

## 2.2. Application of the DL Techniques

### 2.2.1. CLM for Lung Field Defects

On CXR examination, the presence or absence of defects in the lung fields should be considered as the first priority to determine whether the images need to be retaken. This is because if the lung fields to be examined are not adequately depicted without any defects, there is a risk of missing the disease to be detected.

We used 1000 of 4809 images in which the lung field area was depicted without defects (defined as “OK”) and 1000 images in which the lung field area was depicted with defects (defined as “NG”) (Figure 1) and created a CLM for the lung field defects.

**Figure 1.** Representative examples with and without lung field defect.

When this CLM was trained, for data augmentation, the brightness was changed in five steps (50%, 75%, 100%, 125%, and 150%) and the scale was changed in three steps ( $0.8\times$ ,  $0.9\times$ , and  $1.0\times$ ). Therefore, the data were expanded 15 times using data augmentation and used for training to create a CLM.

In this study, data augmentation with scaling was performed to improve the generalization performance. However, if the upscaling process is applied to the images without lung field defects, those images may be the images with lung field defects. Therefore, the equalizing and downscaling processes were applied. In addition, to standardize the data augmentation method for the images with lung field defects, the data were expanded by applying the same method.

### 2.2.2. CLM for Obstacle Shadow

If no lung field defect is found on the CXR image, it is important to check the presence or absence of obstacle shadows and their types to determine whether retaking the images is necessary. This is because the presence of obstacle shadows may lead to overlooking the disease if it is hidden in the location of the obstacle shadows.

There are two major types of obstacle shadows: medical and nonmedical devices. In the case of medical devices, such as pacemakers and tubes, there is no need to remove them during imaging because their removal poses a risk to the patient’s safety. On the other hand, in the case of nonmedical devices, such as accessories and underwear, their removal does not pose a risk to the patient’s body; therefore, these should be removed before retaking the CXR image.

Of the collected 4809 CXR Images, a total of 1000 chest images without any obstacle shadows (defined as “None”); 1000 chest images with medical devices, such as a pacemaker, port, and tube and an electrocardiogram device (defined as “Medical devices”); and 1000 chest images with nonmedical devices, such as a belt, an underwear, a necklace, and earrings (defined as “Others”), were selected to create a CML of the disorder shadows (Figure 2).

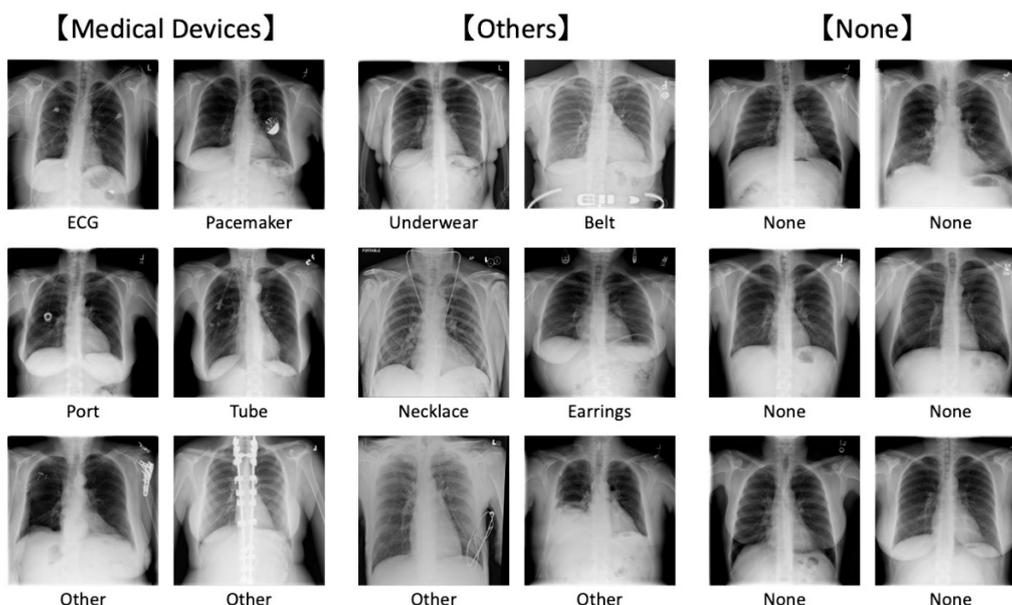


Figure 2. Representative examples with and without obstacle shadow.

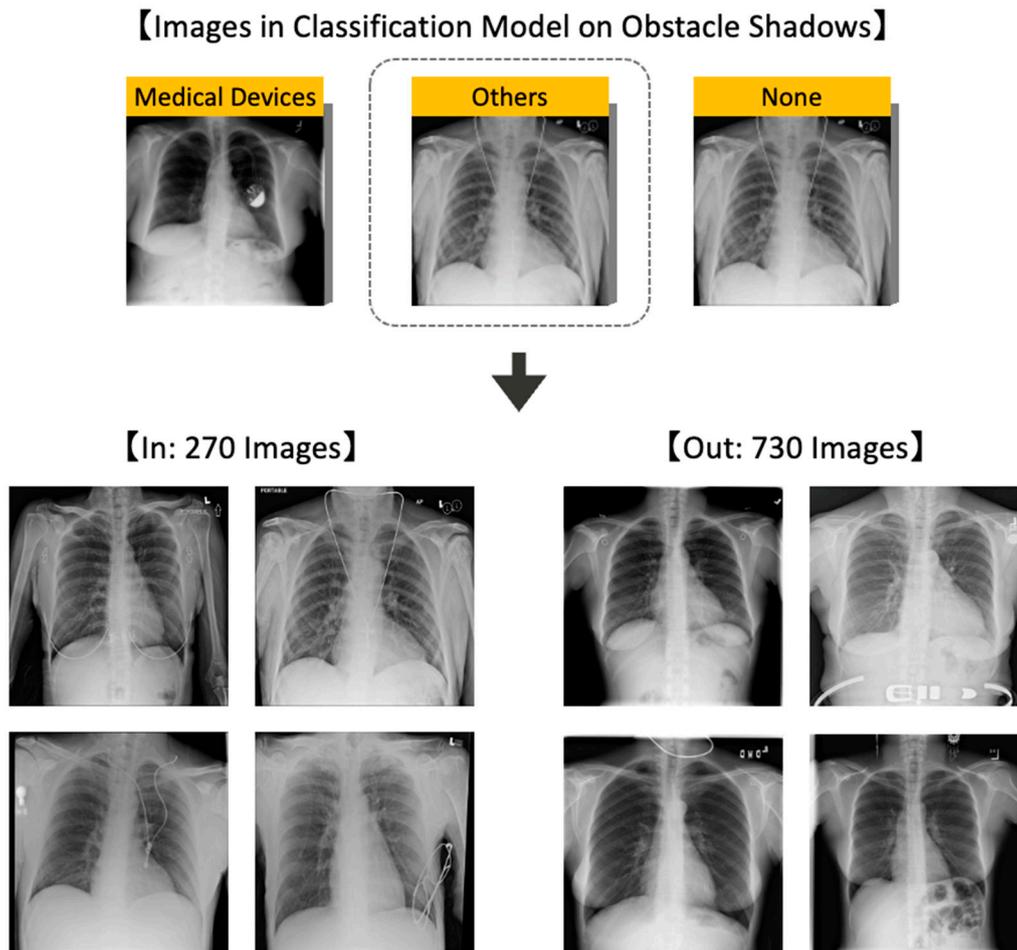
In addition, when this CLM was trained, for data augmentation, the brightness was changed in five steps (50%, 75%, 100%, 125%, and 150%) and the angle was changed in seven steps (−15°, −10°, −5°, 0°, +5°, +10°, and +15°). The training data were multiplied by 35 to create a CLM. Patients with medical devices often undergo portable chest radiography while lying in a hospital bed because of their health conditions. In portable chest radiography, the detector is placed between the bed and the patient for imaging, and the position of the patient may be tilted in relation to the image. Therefore, we decided to use the data augmentation method via rotation to improve the generalization performance to the patient’s body inclination.

### 2.2.3. CLM for the Location of Obstacle Shadow

For CXR images that belonged to the “Other” category in the dataset of CLM for obstacle shadow, the location of the obstacle shadow, inside or outside the lung field, was important to determine the need for a retake the CXR examination. This was because if the obstacle shadows were located inside of the lung fields, the disease may be overlooked,

whereas if they were located outside of the lung fields, they would not affect the diagnosis because the lung fields can be adequately evaluated.

We manually classified 1000 chest images (“Others”) in which a nonmedical device was detected according to the position of the observed obstacle shadows and created two data groups: one group in which the obstacle shadows were located inside of the lung fields (“In”) and the other group in which the obstacle shadows were located outside of the lung fields (“Out”). The numbers of images for the “In” and “Out” data groups were 270 and 730, respectively (Figure 3).



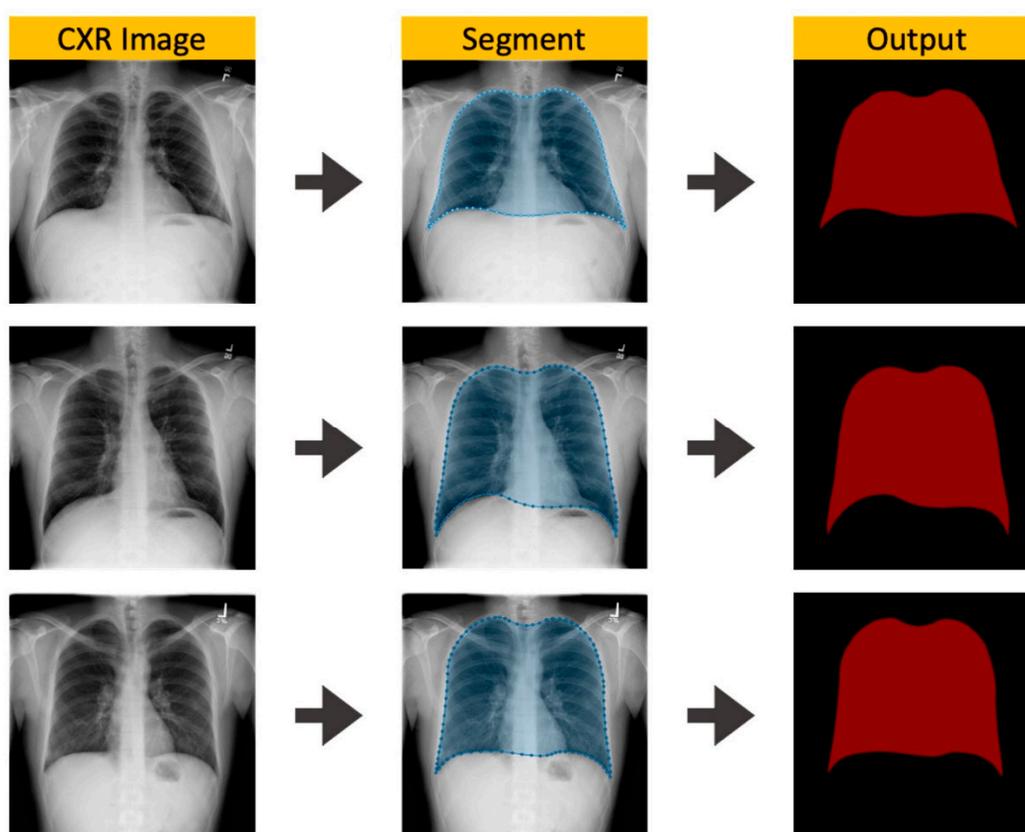
**Figure 3.** Classification of the location of the obstacle shadows.

When this CLM was trained, the brightness change in five steps (50%, 75%, 100%, 125%, and 150%) was used as a common data augmentation method. To maintain the number of training images close to the same number, the “In” image was rotated 11 times ( $-15^\circ$ ,  $-12^\circ$ ,  $-9^\circ$ ,  $-6^\circ$ ,  $-3^\circ$ ,  $0^\circ$ ,  $+3^\circ$ ,  $+6^\circ$ ,  $+9^\circ$ ,  $+12^\circ$ , and  $+15^\circ$ ) and scaled equally 2 times ( $0.9\times$  and  $1.0\times$ ). On the other hand, the “Out” image was rotated 7 times ( $-15^\circ$ ,  $-10^\circ$ ,  $-5^\circ$ ,  $0^\circ$ ,  $+5^\circ$ ,  $+10^\circ$ , and  $+15^\circ$ ) to increase the data volume by a total of 35 times, and then a CLM was created.

#### 2.2.4. SSM for the Lung Field Region

If the human eyes can immediately recognize the lung field region when applying these CLMs, the judgments presented by the created CLMs can be easily recognized. This will contribute to the improvement not only in the throughput of daily work but also in the accuracy of examinations because it will lead to the establishment of a system of second checks by artificial intelligence and humans.

In this study, we performed a segmentation of both lung fields, including the mediastinum, on 1000 chest images and constructed an SSM that recognized the lung field region (Figure 4). The lung fields, including the mediastinum, were segmented because medical devices, such as corrective fixation devices for scoliosis, and nonmedical devices, such as necklaces, were often located on the mediastinum. In other words, in this study, the lung field region, including the mediastinum, underwent semantic segmentation so that these devices can be accurately segmented, including those on the mediastinum.



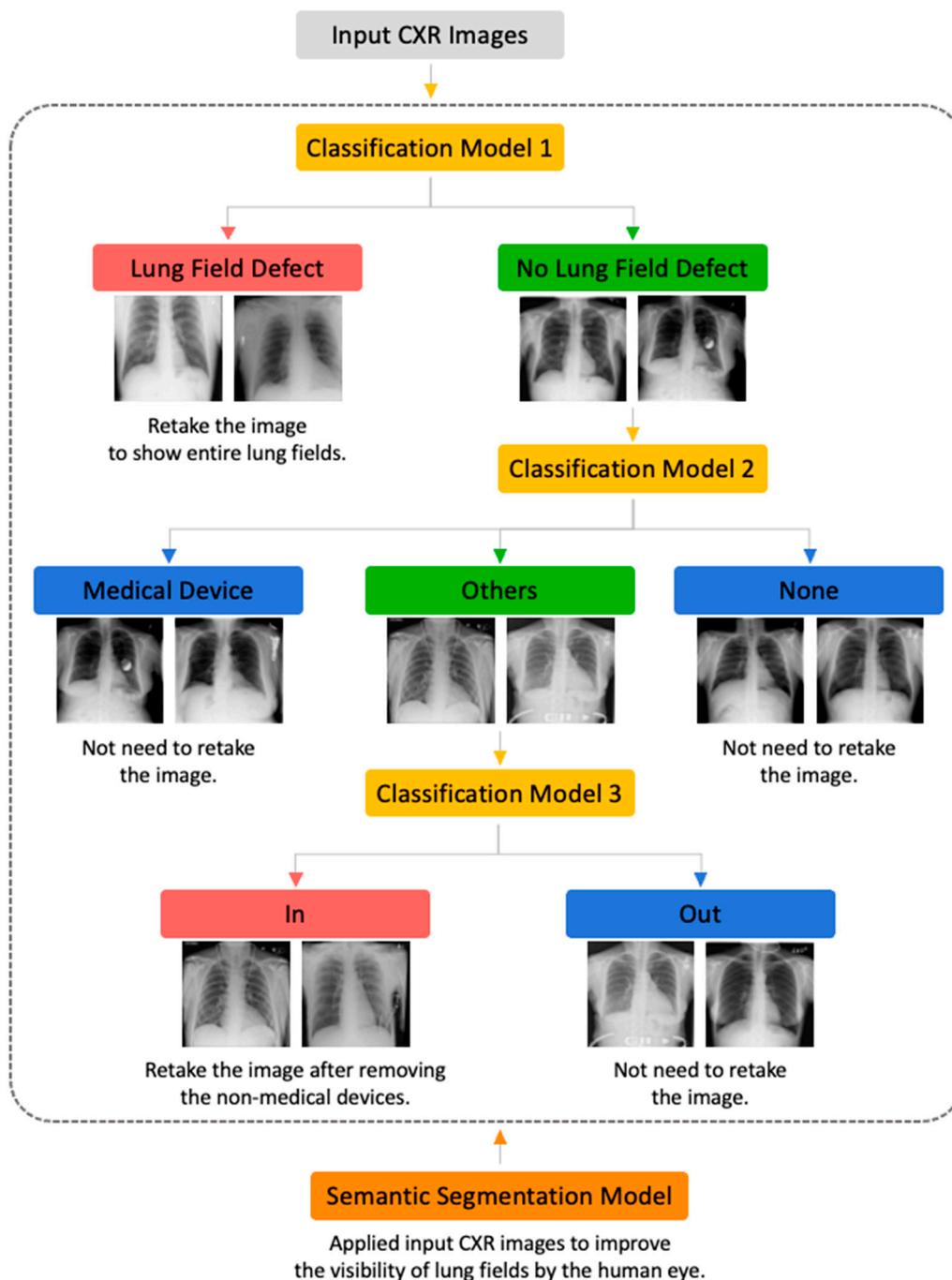
**Figure 4.** Representative examples of segmentation of both lung fields including the mediastinum.

The model was created by training the data augmented a total of 35 times by changing the brightness five times (50%, 75%, 100%, 125%, and 150%) and rotating the data seven times ( $-15^\circ$ ,  $-10^\circ$ ,  $-5^\circ$ ,  $0^\circ$ ,  $+5^\circ$ ,  $+10^\circ$ , and  $+15^\circ$ ).

#### 2.2.5. Software Development

In this study, we developed an in-house MATLAB (The MathWorks, Inc., Natick, MA, USA) software to evaluate the CXR images by combining three CLMs and an SSM in a single software package. With this combination of 4 DL models, the software could not only immediately identify the lung fields (contouring) in the input CXR image but could also obtain an artificial intelligence response from the viewpoint of whether retaking the CXR image is necessary. The definition of the classification model is based on the procedure actually used by radiological technologists to check CXR images taken in clinical practice. Decision making is performed to ensure that all lung fields are included in the image, that no foreign objects are reflected in the image, and that if foreign objects are included, they do not affect the lung fields as the obstacle shadow, including the longitudinal regions. The developed software allows the same flow as that used by radiological technologists to determine whether or not a CXR image needs to be retaken.

Figure 5 shows an overview of our software using multiple DL models.



**Figure 5.** Overview of the CXR image evaluation software.

### 2.3. Architecture and Training

In this study, three CLMs and an SSM were made—ResNet-50 [19] was used to create the CLM, and DeepLabv3+ [20] was used to create the SSM. Tables 2 and 3 show the training conditions used for the CLM and SSM. Table 4 shows the equipment and software used in the training.

To avoid overtraining, which would make it impossible to accurately evaluate the generalization performance, we used a fivefold cross validation for training and evaluation. Each dataset was defined with a name from 1-fold to 5-fold, respectively, to distinguish between datasets. Because of the many types of obstacle shadows, when creating the CLM for the obstacle shadows, datasets were created while taking care not to bias the types of obstacle shadows depicted among the subsets.

**Table 2.** Learning conditions for creating the classification models.

Label	Content
Network	ResNet-50
Minibatch size	256
Epoch	5
Optimizer	Stochastic gradient descent with momentum
Momentum	0.9
Initial learning rate	0.0001
L2 regularization	0.0001

**Table 3.** Training conditions for creating the semantic segmentation model.

Label	Content
Network	DeepLabv3+
Minibatch size	128
Epoch	3
Optimizer	Stochastic gradient descent with momentum
Momentum	0.9
Learn rate drop factor	0.3
Initial learning rate	0.001
L2 regularization	0.005

**Table 4.** Equipment and software used.

Environment	Content
Software	MATLAB 2021a (developed by MathWorks)
OS	Windows 10
CPU	Intel Core i9-10980XE 3.0 GHz
GPU	NDIVIA RTX A6000 48 GB × 2
Memory	DDR4 2933 Quad-Channel

#### 2.4. Evaluating the Created Models

In this study, we evaluated the CLM using overall accuracy on the confusion matrix and the SSM using the mean intersection over union (mIoU). When evaluating “the chest X-ray image evaluation software” with three CLMs and one SSM, the time taken for the input CXR image to pass through each created model and that for the system to show a response as to whether a retaking is necessary (response) were evaluated using the response time on this software (RT). Because each model was created independently, the RT per image of the software was calculated by summing up the RT per image of each model. The formulas for the evaluation indices were as follows, where TP is true positive, FP is false positive, TN is true negative, FN is false negative, N is the number of images,  $RT_{CLM1}$  is the response time of the lung field defect CLM,  $RT_{CLM2}$  is the response time of the disorder shade CLM,  $RT_{CLM3}$  is the response time of the disordered shadow location CLM, and  $RT_{SSM}$  is the response time of the lung field area SSM:

$$Overlay\ Accuracy\ [\%] = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (1)$$

$$mIoU = \frac{1}{N} \sum_{N=1}^N \frac{TP_N}{TP_N + FP_N + FN_N} \quad (2)$$

$$Response\ Time\ [s] = RT_{CLM1} + RT_{CLM2} + RT_{CLM3} + RT_{SSM} \quad (3)$$

Because each model was trained and evaluated using a fivefold cross validation, the mean values of the overall accuracy and mIoU of the obtained five models were used to calculate the evaluation of each model.

### 3. Results

#### 3.1. CLM for Lung Field Defects

Figure 6 and Table 5 show the results of the fivefold cross validation of 1000 chest images without lung field defects and 1000 chest images with lung field defects. Figure 6 shows the results of the model (1-fold) among the five models (1-fold–5-fold), using the confusion matrix as a representative.

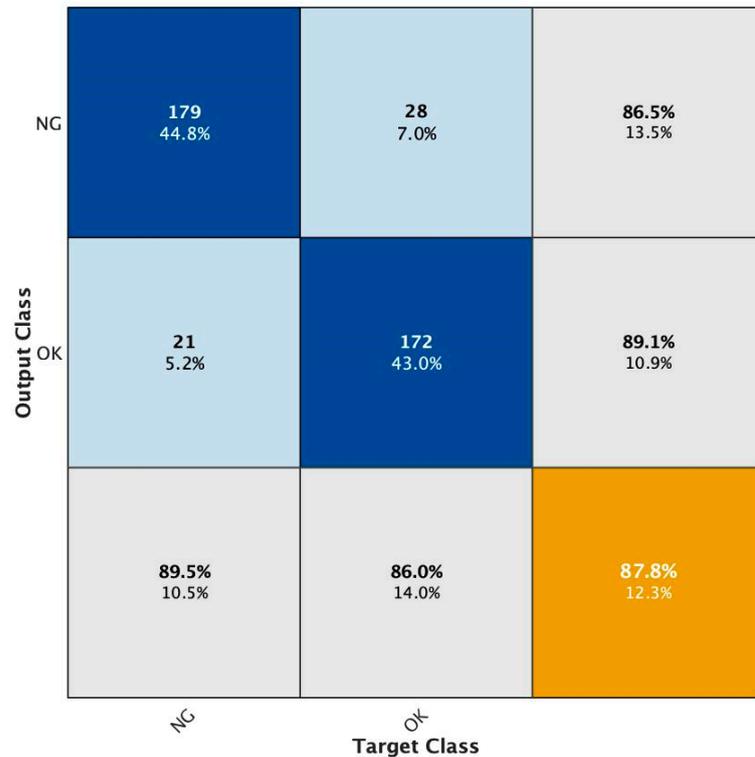


Figure 6. Confusion matrix in dataset (1-fold).

Table 5. Accuracy evaluation for each dataset.

Model	Overall Accuracy (%)
1-fold	87.8
2-fold	89.8
3-fold	92.2
4-fold	89.5
5-fold	89.8
Mean	89.8

These results indicated that the presence or absence of lung field defects could be discriminated with an accuracy of approximately 90%.

#### 3.2. CLM for Obstacle Shadow

Figure 7 and Table 6 show the results of the fivefold cross validation of 1000 chest images with a medical device (“Medical device”), 1000 chest images with the nonmedical device (“Other”), and 1000 chest images without obstacle shadows (“None”). Figure 7 shows the results of model (1-fold) among the five models (1-fold–5-fold), using the confusion matrix as a representative of the results of model (1-fold).

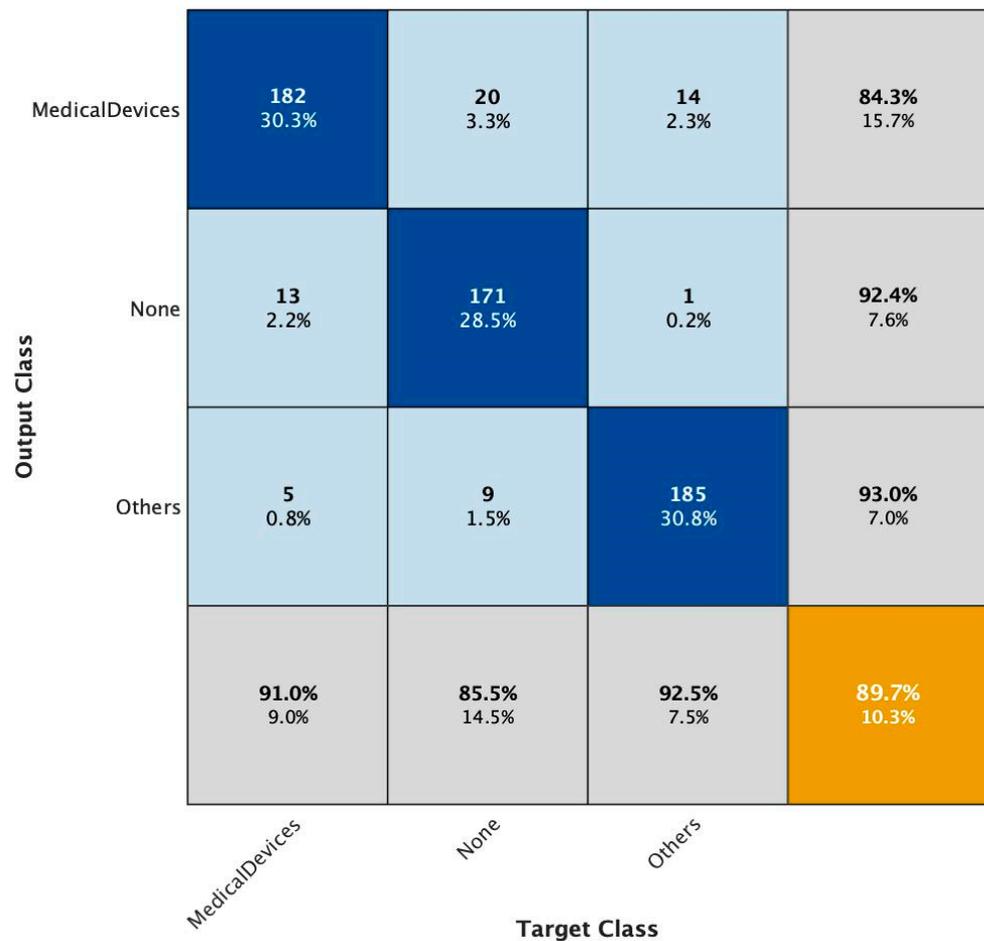


Figure 7. Confusion matrix in dataset (1-fold).

Table 6. Accuracy evaluation for each dataset.

Model	Overall Accuracy (%)
1-fold	89.7
2-fold	90.3
3-fold	93.2
4-fold	93.0
5-fold	92.3
Mean	91.7

These results indicated that the presence or absence and type of the obstacle shadow could be discriminated with accuracy of approximately 92%.

### 3.3. CLM for Location of Obstacle Shadow

Figure 8 and Table 7 show the results of the evaluation of 730 “In” and “Out” images, which were classified according to the location of the obstacle shadows, by the fivefold cross validation. Figure 8 shows the results of model (1-fold) among the five models (1-fold–5-fold), using the confusion matrix as a representative of the results of model (1-fold).

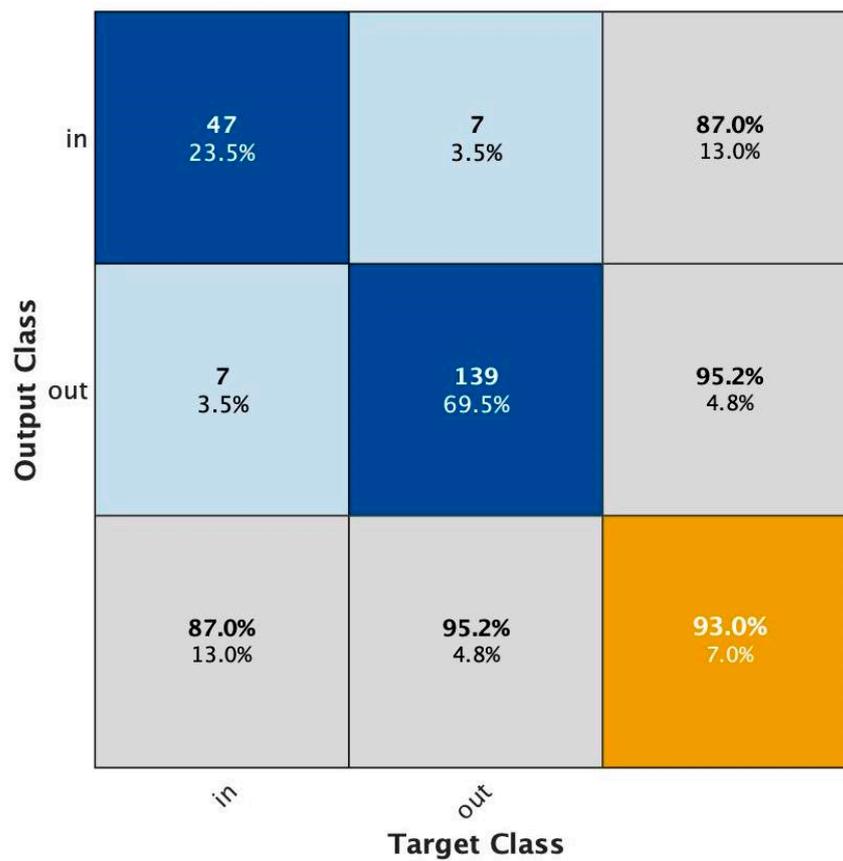


Figure 8. Confusion matrix in dataset of 1-fold.

Table 7. Accuracy evaluation for each dataset.

Model	Overall Accuracy (%)
1-fold	93.0
2-fold	90.0
3-fold	93.5
4-fold	86.5
5-fold	93.0
Mean	91.2

These results indicated that it was possible to discriminate with approximately 91% accuracy whether the location of the obstacle shadow was inside or outside the lung field.

### 3.4. SSM for the Lung Field Region

Semantic segmentation of the lung field region was performed on 1000 chest images, and the results evaluated by the fivefold cross validation are shown in Table 8.

Table 8. Accuracy evaluation for each dataset.

Model	mIoU
1-fold	0.930
2-fold	0.923
3-fold	0.919
4-fold	0.924
5-fold	0.903
Mean	0.920

The results show that semantic segmentation can be performed with 92% accuracy for the lung field region, including the mediastinum.

### 3.5. Chest X-ray Image Evaluation Software

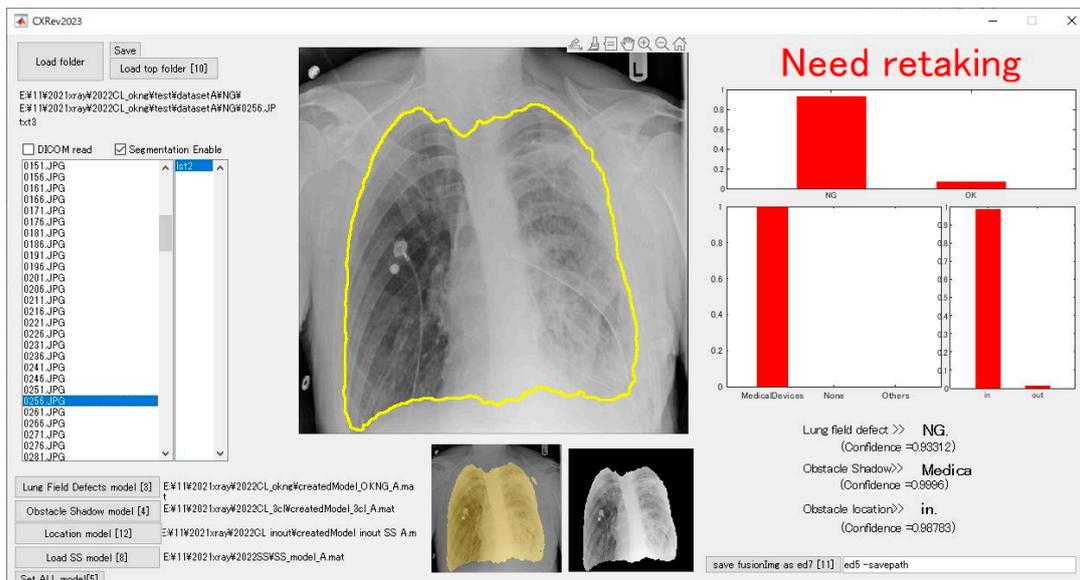
The RT per image for each model was used to evaluate the CXR image evaluation software using three CLMs and one SSM. The results are shown in Table 9.

**Table 9.** Response time of each model on this software (CLM1—lung field defect classification model; CLM2—obstacle shadow classification model; CLM3—obstacle shadow location classification model; and SSM—lung field region semantic segmentation model).

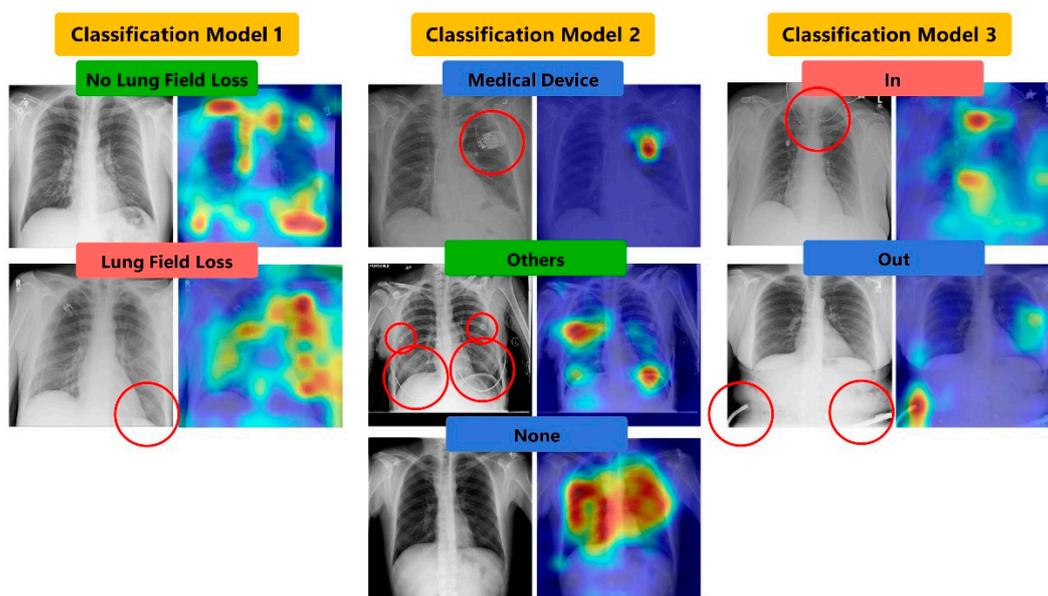
Model	CLM1 (s)	CLM2 (s)	CLM3 (s)	SSM (s)
1-fold	0.95959	1.1518	0.71392	5.1727
2-fold	0.98751	1.1154	0.60750	5.9433
3-fold	1.0289	1.0620	0.59940	5.7609
4-fold	0.89772	1.1368	0.61593	5.9016
5-fold	0.93638	1.0817	0.71462	6.1392
Mean	0.96000	1.1100	0.65000	5.7800

The results were as follows: the RT per chest image for the lung field defect CLM was  $2.40 \times 10^{-3}$  s; that for the obstacle shadow CLM was  $1.85 \times 10^{-3}$  s; that for the obstacle shadow location CLM was  $3.25 \times 10^{-3}$  s; and that for the lung field region SSM was  $2.89 \times 10^{-2}$  s. Therefore, the RT of this software is the sum of these RTs  $3.64 \times 10^{-2}$  s.

Figure 9 shows the actual screen of the software. For example, the response “Need retaking” was shown for the chest image with the lung field defect, indicating the radiographer’s need for a retake. Figure 10 shows the decision evidence for each of the representative images, with a heatmap showing the basis for each decision in terms of occlusion sensitivity [21], which is one of the Saliency maps. For each feature indicated by a red circle on the image, the heatmap shows strong evidence for classification in red.



**Figure 9.** Actual screen of the CXR image evaluation software.



**Figure 10.** Classification evidence for each representative image: each image (left) and heatmap by occlusion sensitivity (right). Red circles: features in the image.

#### 4. Discussion

To the best of our knowledge, this software is the first attempt to apply the DL techniques to determine whether retaking a CXR image is necessary. In this software, four DL models were combined to form a single system. However, since there are few related previous studies, the models and software developed in this study will be objectively discussed through comparison with studies of QA systems for X-ray images, which are most relevant to this study, and studies on semantic segmentation of lung field regions not including the mediastinum. Here, we discussed each model and this software and described the limitations and prospects of this study.

First, we presented a discussion on the lung field defect, CLM. Although the classification technique has been widely applied to the classification of the presence or absence of defective products on factory production lines [22], its application to the classification of defects in the lung field region has not been confirmed yet. In such a situation, Junhao et al. [17] classified the presence or absence of defects in the lung field region as part of the construction of a QA system for CXR images, using a combined application of the semantic segmentation and classification techniques. In their study, they attempted to perform QA pixel-wise using the semantic segmentation technique instead of applying image classification directly, because the target area for QA evaluation was small. As a result, the accuracy of the image classification for the presence or absence of lung field defects at the image level was 92.50%, which was slightly higher than that of the present study, whereas the pixel-wise examination using the semantic segmentation technique showed an accuracy of 97.96%. In the data used in this study, there were some images in which the lung field defects were identified in small areas, such as the costophrenic angle. Considering the performance of hardware in actual clinical practice, complex data input may result in lower throughput due to slower processing speed. However, it is important to consider input features that will allow AI to identify the presence or absence of lung field defects more easily in the future. Second, we presented a discussion on the obstacle shadow CLM. For the classification of the presence/absence of obstacle shadows, Junhao et al. [17] attempted to classify the presence/absence of artifacts in the same way. In this case, the accuracy of the image-level classification of the presence/absence of artifacts was 83.75%, which was slightly lower than the accuracy of 91.7% in the present study. One reason for this difference in accuracy may be the difference in the number of chest images with artifacts used for training. The fact that the accuracy was better in the present study,

in which a larger number of chest images with artifacts were used for training, suggests that further improvement in classification accuracy can be expected with an increase in the number of images in the future. On the other hand, the classification using the semantic segmentation technique showed an accuracy of 94.90%, which was better than the accuracy of the present study.

Although we were unable to identify any studies that directly classified the types of disability shadows, Ue-Hwan et al. [23] investigated the manufacturer's classification, model group identification, and magnetic resonance imaging safety characterization of cardiac implantable electronic devices (CIEDs) using a DL-based algorithm. The overall accuracy rates against each internal test dataset were 99.7%, 97.2%, and 98.9%, respectively. In this study, images with the CIED portion cropped and resized as a preprocess were used for training and evaluation. Therefore, considering the purpose of the study and the difference in image formats, it is difficult to directly compare the accuracy with the present study. Third, we described a discussion of the obstacle shadow location CLM. In this study, we applied the data augmentation technique to 270 chest images that belonged to the "In" category and 730 chest images that belonged to the "Out" category. The number of data used in the training was made to be close to the same number by adjusting the expansion rate of the data augmentation. As a result, although there were some differences in the number of training data between the two classes, the classification accuracy was high without being attracted to the features of the other. This result indicates the validity of the data augmentation method used to maintain the number of data close to the same number. However, one of the reasons why the overall accuracy of the present obstacle shadow location CLM was only 91.2% may be due to the cases in which the obstacle shadows were located at the boundaries of the lung fields or where the obstacle shadows straddled the boundaries between the inside and outside of the boundaries, such as necklaces. In such cases, it is difficult to determine which region of the lung field the obstacle shadow is located, whether inside or outside. Therefore, it is considered that such a situation may have caused the accuracy degradation; however, we have not been able to visualize the basis for the decision of the DL model in this study. In the future, we believe that it is necessary to incorporate a system, such as the saliency maps [21,24], that can represent the basis of judgments on a heat map and examine the causes in more detail. Fourth, we described a discussion on the SSM of the lung field region. In this study, we compared the results of previous studies [11,12] and performed semantic segmentation for the lung field region, including the mediastinum. Among such previous studies, in the study on CardioNet by Abbas et al. [25], semantic segmentation was performed not only for the lung field but also for the heart and clavicle. Among the CardioNet used in this study, CardioNet-B performed semantic segmentation for the lung, heart, and clavicle with mIoU values of 0.9728, 0.9042, and 0.8674, respectively. Thus, the accuracy was higher than that of the present study when the comparison was made only for the lung. However, when we focused on the mIoU values of the heart and clavicle, we could confirm that the mIoU values were lower than those of the lung. This result suggests that semantic segmentation is more difficult in the low-radiolucent tissues than in the high-radiolucent tissues, such as the lung field. This also suggests that the accuracy of semantic segmentation of the lung field, including the mediastinum, tends to be lower than that of the semantic segmentation of the lung field without the mediastinum.

However, the SSM developed in this study had a higher mIoU value than in a previous study [15] using DeepLabv3+, similar to our method. Considering this finding, we believe that the semantic segmentation and data augmentation methods used in our method are appropriate. However, to further improve the accuracy of the semantic segmentation method in the future, it is natural to use more data, and it is necessary to take measures proposed by Johnatan et al. [26] to reduce the accuracy of semantic segmentation due to the abnormal shadows of lesions and accurately recognize the lung field regions of more examinees. Fifth, we discussed the evaluation software for CXR images that combined these DL models. We evaluated the RT per chest image of the software by summing the

RTs per chest image of the four DL models. Considering that the image processing time of FPD was approximately several seconds and that of CR was approximately several tens of seconds, the RT of  $3.64 \times 10^{-2}$  s of this software is almost the same as the time required for conventional imaging operations, and it is considered to be able to provide artificial intelligence's judgment on whether retaking is necessary. However, because the RT of this software varies depending on the device used, it is important to examine the response time on a PC with the specifications used in actual clinical practice in the future. In addition, the present study did not evaluate FLOPs. Only the response time on this software was calculated as a result, which should be taken into consideration in the future. FLOPs are useful for evaluating the performance and efficiency of models [27], but in this study, we were interested in evaluating the time from image input to the display of results on a simple software program. This is because the actual time to display multiple models in software is one of the criteria for clinical image confirmation.

This software was created by combining multiple DL models. All images in this study included images from healthy patients to patients who were hospitalized and placed under long-term management, including electrocardiograms. Typically, inappropriate X-ray images depicting missing lung fields are retaken, so they rarely remain in the data. The reason for multiple segmentations and classifiers is that a defective lung field is not appropriate for medical imaging when detecting a lesion, and the radiological technologist must take this into consideration. Therefore, the model should focus only on the detection of lung field defects. In addition, obstruction shadows can be allowed or not, depending on their location, so the segmentation model of the lung field, including the mediastinum, needs to show the exact lung field. Therefore, in this study, the software was developed based on the radiological technologist's decision in X-ray radiography. Each model was created by learning the images that were originally subject to retaking and could not be stored in the picture archiving and communication system or those that were drawn by a medical device because of health conditions. Therefore, for the effective use of the original images, each model was created independently without considering the overlap of training data. This suggests that it is difficult to comprehensively evaluate the software that combines each model. However, because each model of this software had an accuracy of approximately 90%, it is considered to be possible to immediately accurately provide the radiologist with a decision as to whether the retaking is necessary, and to encourage confirmation by the human eye. Finally, the limitations and prospects of this study were described. One of the limitations of this study is that the software was built and evaluated using CXR images collected from "CXR8" published by the NIH Clinical Center. Considering that CXR examination is the most common imaging method, it is important for a QA-related system, such as this software, to be trained and evaluated on data taken from several facilities. This is because it contributes to higher generalization performance. Therefore, we will aim to add more data by utilizing other datasets in the future and attempt to construct software with high generalization performance and accuracy. Even though the software developed in this study works on MATLAB, the models created in MATLAB can be converted to the open neural network exchange (ONNX) format, which allows for improvements and refinements irrespective of the development framework. Another limitation is that the contents considered in this system alone cannot provide appropriate judgments for all cases. For example, we were unable to examine the effects of the scapula on the lung field, as examined by Junhao et al. [17]. Therefore, additional data collection and training will be necessary to apply the method to more situations. In addition, the efficiency of post-imaging should be considered for the improvement in the overall efficiency of CXR examinations. Oura et al. [28] reported the QA of CXR images using the DL techniques. In this study, the DL techniques were applied to four points—correction of orientation, correction of angle, correction of left–right reversal, and judging the patient's position—and proposed a method to improve the accuracy and efficiency of daily operations. Therefore, the combined application with other QA systems in the future will enable us to perform the current CXR examinations with higher throughput. For example, a study that developed

a computer-aided diagnosis system (CAD system) that utilizes a convolutional neural network ensemble to reduce the workload of physicians and radiologists and achieve quick and accurate diagnosis showed a marked improvement in accuracy in the classification of chest X-ray images [29]. In addition, a study that developed a DL-based algorithm to reduce data acquisition time in 3-D X-ray microscopy showed an 8- to 10-fold increase in speed while maintaining image quality, even with several hundred X-ray projections [30]. There are studies that have achieved high accuracy for each of these purposes, and further studies are needed in the system to determine whether or not to perform retakes in this study. Since the penetration of X-ray images varies depending on the subject's physique and other factors, it is necessary to be able to automatically control image quality and detect low image quality to improve the efficiency of medical image analysis, as in the system developed by Dovganich et al. [31] to automatically determine the penetration of pulmonary X-ray images. In addition, software and hardware and their efficiency need to be considered for future applications of inference; further study on how to make classification and regression models in the development of deep learning models, as shown by Sumathi et al. [32]; and for early prediction of infectious diseases based on chest X-ray images, as shown by Namburu et al. [33] and the integration of deep learning algorithms with FPGA hardware for efficient analysis and low power consumption, among other improvements. Based on these improvements, further development of the method in this study is considered possible in the future.

## 5. Conclusions

In this study, we constructed software to determine whether a CXR image needs to be retaken by combining three CLMs and an SSM. In addition to the 90% accuracy of each model, we found that the software can provide a decision on whether a retaking is necessary in  $3.64 \times 10^{-2}$  s. The developed software can provide the same decision-making support that a radiologist would require to determine whether a CXR image needs to be retaken, and further improvements to the individual classification models are expected to lead to a system more appropriate for real-world clinical practice. However, there is a need to further investigate developmental considerations by incorporating existing techniques that have been presented in many fields.

**Author Contributions:** K.U. contributed to the data analysis, algorithm construction, and writing and editing of the manuscript. T.Y. and S.I. reviewed and edited the manuscript. H.S. proposed the idea and contributed to the data acquisition, performed supervision and project administration, and reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The created models in this study are available on request from the corresponding author. The source code of this study is available at <https://github.com/MIA-laboratory/CXRevaluation> (accessed on 20 April 2023).

**Acknowledgments:** The authors would like to thank the laboratory members of the Medical Image Analysis Laboratory for their help.

**Conflicts of Interest:** The authors declare that no conflict of interest exist.

## References

1. Sharma, R.; Sharma, S.D.; Pawar, S.; Chaubey, A.; Kantharia, S.; Babu, D.A.R. Radiation Dose to Patients from X-ray Radiographic Examinations Using Computed Radiography Imaging System. *J. Med. Phys. Assoc. Med. Phys. India* **2015**, *40*, 29–37. [[CrossRef](#)] [[PubMed](#)]
2. Moore, C.S.; Wood, T.J.; Saunderson, J.R.; Beavis, A.W. Correlation of the Clinical and Physical Image Quality in Chest Radiography for Average Adults with a Computed Radiography Imaging System. *Br. J. Radiol.* **2013**, *86*, 20130077. [[CrossRef](#)] [[PubMed](#)]

3. Lin, C.S.; Chan, P.C.; Huang, K.H.; Lu, C.F.; Chen, Y.F.; Lin Chen, Y.O. Guidelines for Reducing Image Retakes of General Digital Radiography. *Adv. Mech. Eng.* **2016**, *8*, 1–6. [[CrossRef](#)]
4. Yoshimura, T.; Nishioka, K.; Hashimoto, T.; Mori, T.; Kogame, S.; Seki, K.; Sugimori, H.; Yamashina, H.; Nomura, Y.; Kato, F.; et al. Prostatic Urinary Tract Visualization with Super-Resolution Deep Learning Models. *PLoS ONE* **2023**, *18*, e0280076. [[CrossRef](#)] [[PubMed](#)]
5. Ichikawa, S.; Hamada, M.; Sugimori, H. A Deep-Learning Method Using Computed Tomography Scout Images for Estimating Patient Body Weight. *Sci. Rep.* **2021**, *11*, 15627. [[CrossRef](#)]
6. Manabe, K.; Asami, Y.; Yamada, T.; Sugimori, H. Improvement in the Convolutional Neural Network for Computed Tomography Images. *Appl. Sci.* **2021**, *11*, 1505. [[CrossRef](#)]
7. Sugimori, H.; Hamaguchi, H.; Fujiwara, T.; Ishizaka, K. Classification of Type of Brain Magnetic Resonance Images with Deep Learning Technique. *Magn. Reson. Imaging* **2021**, *77*, 180–185. [[CrossRef](#)]
8. Asami, Y.; Yoshimura, T.; Manabe, K.; Yamada, T.; Sugimori, H. Development of Detection and Volumetric Methods for the Triceps of the Lower Leg Using Magnetic Resonance Images with Deep Learning. *Appl. Sci.* **2021**, *11*, 12006. [[CrossRef](#)]
9. Sugimori, H.; Sugiyama, T.; Nakayama, N.; Yamashita, A. Development of a Deep Learning-Based Algorithm to Detect the Distal End of a Surgical Instrument. *Appl. Sci.* **2020**, *10*, 4245. [[CrossRef](#)]
10. Chen, H.; Miao, S.; Xu, D.; Hager, G.D.; Harrison, A.P. Deep Hierarchical Multi-Label Classification Applied to Chest X-ray Abnormality Taxonomies. *Med. Image Anal.* **2020**, *66*, 101811. [[CrossRef](#)]
11. Arvind, S.; Tembhurne, J.V.; Diwan, T.; Sahare, P. Improved Light Weight Deep CNN Based U-Net for the Semantic Segmentation of Lungs from Chest X-rays. *Results Eng.* **2023**, *17*, 100929. [[CrossRef](#)]
12. Liu, W.; Luo, J.; Yang, Y.; Wang, W.; Deng, J.; Yu, L. Automatic Lung Segmentation in Chest X-ray Images Using Improved U-Net. *Sci. Rep.* **2022**, *12*, 8649. [[CrossRef](#)] [[PubMed](#)]
13. Harris, R.J.; Baginski, S.G.; Bronstein, Y.; Kim, S.; Lohr, J.; Towey, S.; Velichkovich, Z.; Kabachenko, T.; Driscoll, I.; Baker, B. Measurement of Endotracheal Tube Positioning on Chest X-ray Using Object Detection. *J. Digit. Imaging* **2021**, *34*, 846–852. [[CrossRef](#)] [[PubMed](#)]
14. Sharma, A.; Rani, S.; Gupta, D. Artificial Intelligence-Based Classification of Chest X-ray Images into COVID-19 and Other Infectious Diseases. *Int. J. Biomed. Imaging* **2020**, *2020*, 8889023. [[CrossRef](#)]
15. Štifanić, D.; Musulin, J.; Jurilj, Z.; Šegota, S.B.; Lorencin, I.; Anđelić, N.; Vlahinić, S.; Blagojević, A.; Filipović, N.; Car, Z. Semantic Segmentation of Chest X-ray Images Based on the Severity of COVID-19 Infected Patients. *EAI Endorsed Trans. Bioeng. Bioinforma.* **2021**, *1*, 1–8. [[CrossRef](#)]
16. Matsumoto, H.; Sasa, T.; Uemura, H.; Kaji, D. *Automatic Detection of DR Images in Which the Lungs Are Partially Indistinguishable and of DR Images That Are Motion Blurred*; Konica Minolta Technology Report; Konica Minolta: Chiyoda, Tokyo, Japan, 2014; Volume 11.
17. Hu, J.; Zhang, C.; Member, S.; Zhou, K. Chest X-ray Diagnostic Quality Assessment: How Much Is Pixel-Wise Supervision Needed? *IEEE Trans. Med. Imaging* **2022**, *41*, 1711–1723. [[CrossRef](#)]
18. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. ChestX-Ray8: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2097–2106. [[CrossRef](#)]
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
20. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818. [[CrossRef](#)]
21. Zeiler, M.D.; Fergus, R. *Visualizing and Understanding Convolutional Networks*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8689, ISBN 9783319105895.
22. Fan, S.; Li, J.; Zhang, Y.; Tian, X.; Wang, Q. On Line Detection of Defective Apples Using Computer Vision System Combined with Deep Learning Methods. *J. Food Eng.* **2020**, *286*, 110102. [[CrossRef](#)]
23. Kim, U.; Kim, M.Y.; Park, E.; Lee, W.; Lim, W.; Kim, H.; Oh, S.; Jin, K.N. Deep Learning-Based Algorithm for the Detection and Characterization of MRI Safety of Cardiac Implantable Electronic Devices on Chest Radiographs. *Korean J. Radiol.* **2021**, *22*, 1918–1928. [[CrossRef](#)]
24. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.
25. Jafar, A.; Hameed, M.T.; Akram, N.; Waqas, U.; Kim, H.S.; Naqvi, R.A. CardioNet: Automatic Semantic Segmentation to Calculate the Cardiothoracic Ratio for Cardiomegaly and Other Chest Diseases. *J. Pers. Med.* **2022**, *12*, 988. [[CrossRef](#)]
26. Souza, J.C.; Otávio, J.; Diniz, B.; Ferreira, J.L.; Lucca, G.; Silva, A.C.; Paiva, A.C. De An Automatic Method for Lung Segmentation and Reconstruction in Chest X-ray Using Deep Neural Networks. *Comput. Methods Programs Biomed.* **2019**, *177*, 285–296. [[CrossRef](#)]
27. Wei, S.; Wang, Z.; Sun, Z.; Liao, F.; Li, Z.; Zou, L.; Mi, H. A Family of Automatic Modulation Classification Models Based on Domain Knowledge for Various Platforms. *Electron.* **2023**, *12*, 1820. [[CrossRef](#)]

28. Oura, D.; Sato, S.; Honma, Y.; Kuwajima, S.; Sugimori, H. Quality Assurance of Chest X-ray Images with a Combination of Deep Learning Methods. *Appl. Sci.* **2023**, *13*, 2067. [[CrossRef](#)]
29. Visuña, L.; Yang, D.; Garcia-Blas, J.; Carretero, J. Computer-Aided Diagnostic for Classifying Chest X-ray Images Using Deep Ensemble Learning. *BMC Med. Imaging* **2022**, *22*, 178. [[CrossRef](#)]
30. Villarraga-Gómez, H.; Norouzi Rad, M.; Andrew, M.; Andreyev, A.; Sanapala, R.; Omlor, L.; Graf vom Hagen, C. Improving Throughput and Image Quality of High-Resolution 3D X-ray Microscopes Using Deep Learning Reconstruction Techniques. In Proceedings of the 11th Conference on Industrial Computed Tomography (iCT), Wels, Austria, 8–11 February 2022; Volume 27, pp. 3–8. [[CrossRef](#)]
31. Dovganich, A.A.; Khvostikov, A.V.; Krylov, A.S.; Parolina, L.E. Automatic Quality Control in Lung X-ray Imaging with Deep Learning. *Comput. Math. Model.* **2021**, *32*, 276–285. [[CrossRef](#)]
32. Sumathi, D.; Alluri, K. Deploying Deep Learning Models for Various Real-Time Applications Using Keras. In *EAI/Springer Innovations in Communication and Computing*; Prakash, K.B., Kannan, R., Alexander, S.A., Kanagachidambaresan, G.R., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 113–143, ISBN 978-3-030-66519-7.
33. Namburu, A.; Sumathi, D.; Raut, R.; Jhaveri, R.H.; Dhanaraj, R.K.; Subbulakshmi, N.; Balusamy, B. FPGA-Based Deep Learning Models for Analysing Corona Using Chest X-ray Images. *Mob. Inf. Syst.* **2022**, *2022*, 2110785. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.