HEALTH CARE SCIENCE

ORIGINAL ARTICLE

Leveraging anatomical constraints with uncertainty for pneumothorax segmentation

Han Yuan[1] | Chuan Hong[2] | Nguyen Tuan Anh Tran[3] | Xinxing Xu[4] | Nan Liu[1,5,6]

[1]Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore

[2]Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, USA

[3]Department of Diagnostic Radiology, Singapore General Hospital, Singapore

[4]Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore

[5]Programme in Health Services and Systems Research, Duke-NUS Medical School, Singapore

[6]Institute of Data Science, National University of Singapore, Singapore

Correspondence
Nan Liu, Centre for Quantitative Medicine, Duke-NUS Medical School, 8 College Rd, 169857, Singapore.
Email: liu.nan@duke-nus.edu.sg

Abstract

Background: Pneumothorax is a medical emergency caused by the abnormal accumulation of air in the pleural space—the potential space between the lungs and chest wall. On 2D chest radiographs, pneumothorax occurs within the thoracic cavity and outside of the mediastinum, and we refer to this area as "lung + space." While deep learning (DL) has increasingly been utilized to segment pneumothorax lesions in chest radiographs, many existing DL models employ an end-to-end approach. These models directly map chest radiographs to clinician-annotated lesion areas, often neglecting the vital domain knowledge that pneumothorax is inherently location-sensitive.

Methods: We propose a novel approach that incorporates the lung + space as a constraint during DL model training for pneumothorax segmentation on 2D chest radiographs. To circumvent the need for additional annotations and to prevent potential label leakage on the target task, our method utilizes external datasets and an auxiliary task of lung segmentation. This approach generates a specific constraint of lung + space for each chest radiograph. Furthermore, we have incorporated a discriminator to eliminate unreliable constraints caused by the domain shift between the auxiliary and target datasets.

Results: Our results demonstrated considerable improvements, with average performance gains of 4.6%, 3.6%, and 3.3% regarding intersection over union, dice similarity coefficient, and Hausdorff distance. These results were consistent across six baseline models built on three architectures (U-Net, LinkNet, or PSPNet) and two backbones (VGG-11 or MobileOne-S0). We further conducted an ablation study to evaluate the contribution of each component in

---

Abbreviations: ACR, American College of Radiology; AUROC, area under the receiver operating characteristic curve; DL, deep learning; DSC, dice similarity coefficient; HD, Hausdorff distance; IoU, intersection over union; JSRT, Japanese Society of Radiological Technology; MC, Montgomery County data set; ML, machine learning; NPV, negative predictive value; PPV, positive predictive value; SGD, stochastic gradient descent; SIIM, Society for Imaging Informatics in Medicine.

Han Yuan and Chuan Hong contributed equally to this study.

----

the proposed method and undertook several robustness studies on hyperparameter selection to validate the stability of our method.

Conclusions: The integration of domain knowledge in DL models for medical applications has often been underemphasized. Our research underscores the significance of incorporating medical domain knowledge about the location-specific nature of pneumothorax to enhance DL-based lesion segmentation and further bolster clinicians' trust in DL tools. Beyond pneumothorax, our approach is promising for other thoracic conditions that possess location-relevant characteristics.

1 | BACKGROUND

1.1 | Motivation

Pneumothorax is a medical emergency caused by the abnormal accumulation of air in the pleural space, which is the potential space between the lungs and chest wall [1, 2]. The pleural space is non-existent in healthy controls, whereas it can occupy significant portions of the thoracic cavity in patients with a large pneumothorax. Chest radiographs serve as the primary diagnostic tool for pneumothorax, aiding in identifying its location and estimating its size [3, 4]. Traditionally, radiologists report chest radiographs based on their domain knowledge and past experience [5, 6]. In recent years, with the advent of machine learning (ML), especially deep learning (DL), there has been a shift towards automated detection and segmentation of pneumothorax from chest radiographs, achieving promising performance when paired with high-quality annotations [7, 8]. Conventional ML methods focus on identifying contrast regions from chest radiographs [9, 10] and performing morphological operations such as atlas guidance [11], contour deformation [10], and inhomogeneity correction [8] to delineate lesion segmentation. Compared with conventional ML methods involving multiple stages that allow for the incorporation of domain knowledge [12, 13], DL methods typically follow the end-to-end paradigm [14] to capture intricate patterns and contextual information in chest radiographs and offer more fine-grained lesion segmentation. However, end-to-end approaches often neglect the fact that pneumothorax is inherently location-sensitive, predominantly manifesting within the pleural space [15, 16]. Our approach aims to leverage the medical domain knowledge of disease occurrence into DL-based pneumothorax segmentation, emphasizing the significance of the location-specific nature of the disease.

1.2 | Related work

1.2.1 | DL-aided diagnosis for pneumothorax

DL-aided tools have emerged as promising solutions for diagnosing pneumothorax from chest radiographs [17]. Numerous studies have reported accurate classification using various DL models [18], such as AlexNet [19, 20], VGG [19, 21], Inception [19, 22], EfficientNet [23, 24], Network in Network [25], ResNet [26, 27], DenseNet [19, 28], AlbuNet [29, 30], Spatial Transformer [31, 32], and others. While most studies employ the conventional end-to-end approach, Chen et al. designed a two-stage model for pneumothorax classification [33], combining an object detection model YOLO [34] and classification techniques ResNet [27] and DenseNet [28] to distinguish pneumothorax patients and healthy individuals based on the cropped lung field.

Recent research has expanded beyond merely image-level labeling, emphasizing pixel-level lesion area delineation. This not only reduces radiologists' workload but also fosters greater trust in automated systems [35, 36]. Like the binary classification of pneumothorax, DL-based techniques are the state-of-the-art choices for pneumothorax area segmentation [37]. The techniques used mainly comprise the architectures of U-Net [38, 39], DeepLab [40, 41], Mask R-CNN [40, 42], with the backbones of ResNet [39, 43], SE-ResNext [40, 43], DenseNet [15, 43], EfficientNet [40]. Apart from tailoring models on specific datasets, commercial systems, such as DEEP: CHEST-XR-03, have also been

developed and clinically validated for pneumothorax segmentation [44].

1.2.2 | Domain knowledge in chest radiograph analysis

While most of the models mentioned above employ an end-to-end approach, mapping input chest radiographs directly to target disease labels or lesion annotations, the integration of domain knowledge in medical image analysis is pivotal [45, 46]. Incorporating such knowledge not only enhances model interpretability but also boosts model performance [47]. In chest radiograph analysis, disease localization serves as invaluable prior knowledge [48]. For example, Li et al. identified specific anatomical regions within the lung zone, highlighting their diagnostic relevance in detecting conditions of cardiomegaly and pleural effusion [49]. Similarly, Crosby et al. focused on the upper third of chest radiographs and demonstrated that a VGG-based classifier trained barely on the sub-region yielded good performance in the task of distinguishing pneumothorax [50]. Recently, Jung et al. proposed to extract the domain knowledge of thoracic disease occurrence from the class activation map of a thoracic disease classifier. The extracted knowledge was then used for precise thoracic disease localization [51]. Such methods typically utilize masks at various stages, from input to output, ensuring the model's attention is directed towards areas with a higher likelihood of disease presence. More recently, Bateson et al. introduced an innovative method that incorporates prior knowledge about organ size into the model training [52, 53]. By penalizing deviations from this domain knowledge, the model is guided to align more closely with clinical insights. Notably, in tasks of spine and heart segmentation, their method achieved considerably better performance than conventional methods.

1.3 | Contribution

According to clinical knowledge [1, 2], pneumothorax is an abnormal gas collection in the pleural space, which is a potential space between the lungs and chest wall. On a 2D projection of a chest radiograph, pneumothorax localizes inside the thoracic cavity and outside of the mediastinum. We refer to this area as "lung + space," which includes the lungs and the pleural space on a chest radiograph. We propose to incorporate the location information into pneumothorax segmentation. Inspired by [54, 55], we employed occurrence information as a guiding constraint in the segmentation training. To circumvent the need for additional domain knowledge annotation, we proposed a four-phase pipeline to determine the disease occurrence area. Leveraging both external open-access datasets and existing annotations, our evaluations, conducted on a widely-recognized pneumothorax segmentation data set, revealed that our constrained training approach consistently outperformed traditional methods across a range of architectures and backbones. Moreover, to validate the efficacy of our approach, we have conducted an ablation study and two robustness experiments. These studies underscored the effectiveness of individual components within our design and affirmed the overall stability of our proposed method. We hope the proposed method provides a useful framework for embedding domain knowledge into the diagnosis of other thoracic conditions that possess location-relevant characteristics.

2 | METHODS

To harness the power of domain knowledge in improving DL efficacy in lesion segmentation, we propose a four-phase pipeline that incorporates disease occurrence knowledge into the training stage of the pneumothorax segmenter. In this section, we first provide an overview of the proposed pipeline. We then detail the generation process of anatomical constraints. Finally, we elaborate on the constrained training strategy for lesion segmentation.

2.1 | Overview of the proposed pipeline

As outlined in Figure 1, the proposed pipeline contains four indispensable phases to obtain sample-specific lung + space, select well-behaved lung + space, and implement constrained training of pneumothorax segmenter. In Phase 1, we develop an auxiliary lung segmenter using three public lung segmentation datasets, including Japanese Society of Radiological Technology data set (JSRT) [56], Shenzhen data set (Shenzhen) [57], and Montgomery County data set (MC) [57]. This segmenter is then integrated with morphological operations, including connected component cutoff, closing, and dilation to derive a lung + space segmenter in Phase 2. This refined segmenter is subsequently deployed on the target data set of pneumothorax segmentation to predict lung + space. In Phase 3, we introduce a lung + space discriminator, crafted using the training and validation data set for the pneumothorax segmentation, filtering out inaccurately predicted lung + spaces from Phase 2, ensuring only high-quality constraints are retained. In Phase 4, with the selected lung + space from Phase 3, we proceed to train
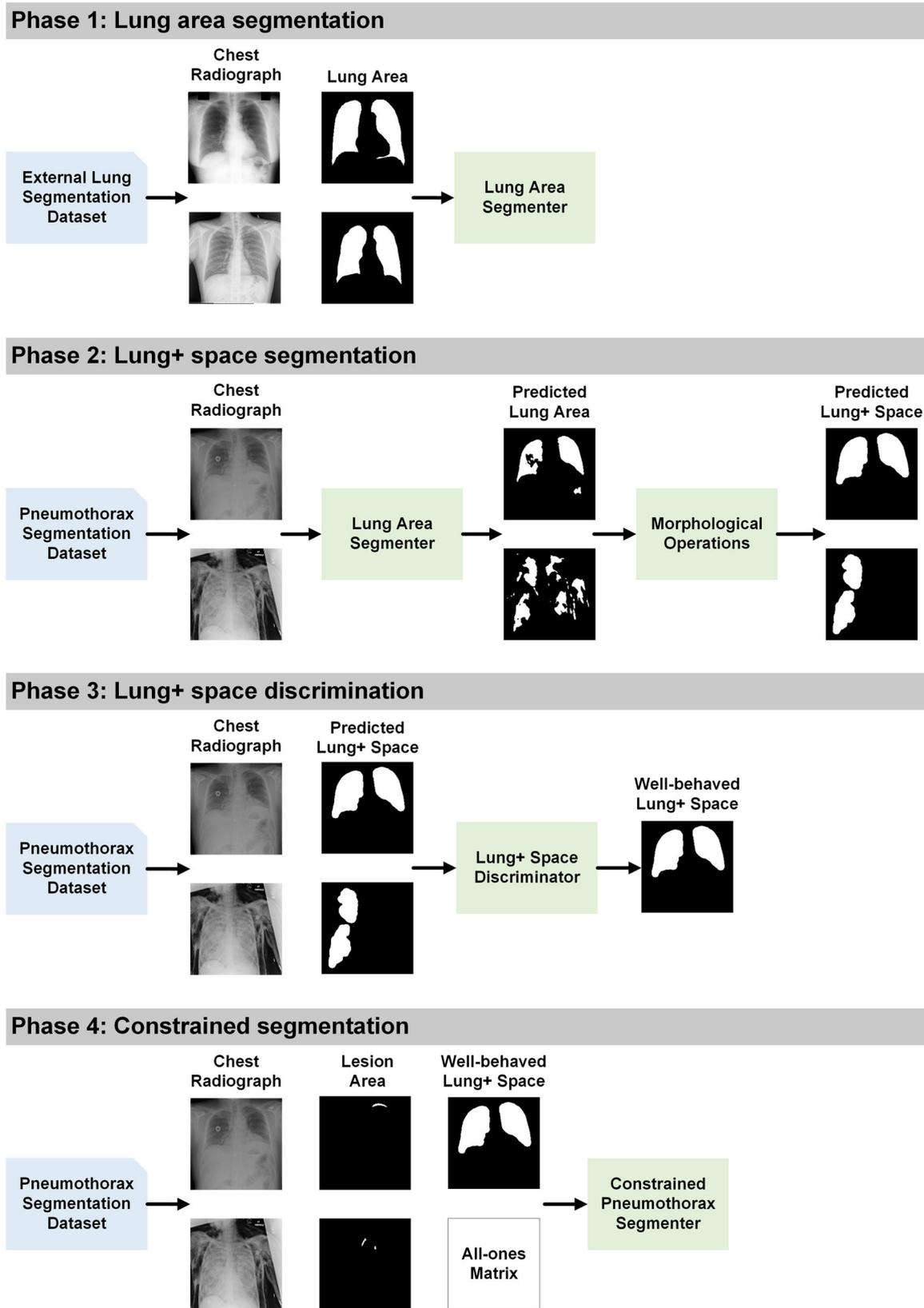
Phase 1: Lung area segmentation

Phase 2: Lung+ space segmentation

Phase 3: Lung+ space discrimination

Phase 4: Constrained segmentation

FIGURE 1 Schematic diagram of the proposed constrained segmentation framework.

the pneumothorax segmenter using a constrained approach. To detail the four phases, the remaining parts are organized as follows: First, we provide in Section 2.2 the design details of Phases 1, 2, and 3 for generating the well-behaved anatomical constraints for pneumothorax, including three coherent modules of lung area segmenter, lung + space segmenter, and lung + space discriminator. Second, we illustrate the formulation of a standard lesion segmenter and outline the proposed constrained segmenter of Phase 4 in Section 2.3.

2.2 | Learning the anatomical constraints

The primary challenge in our proposed method lies in the derivation of well-behaved constraints. From the domain knowledge, we understand that pneumothorax occurs in the lung + space and the lung + space can be used as a constraint in segmenter training. The previous work generally requires constraint-relevant annotation on the target data set, which incurs additional expenses and lacks adaptability when extended to other datasets [54, 58]. To tackle this problem, we exploit three external open-access datasets and an auxiliary task of lung segmentation to generate the lung + space as our anatomical constraints. Specifically, we first generate a lung area segmenter based on the external datasets. Then, we integrate several post-processing techniques into the lung area segmenter to achieve a lung + space segmenter and deploy it on the target data set. Finally, considering the inherent noise in the constraints generated from the external source-based model, a lung + space discriminator is developed to filter out the noisy constraints while retaining the informative ones for the constrained training of the pneumothorax segmenter.

2.2.1 | Phase 1: Lung area segmentation

The foundation for our anatomical constraint learning is the lung area segmenter trained on external datasets. We utilize three publicly available chest radiograph datasets with lung area annotations and develop the lung area segmenter following the standard area segmentation [56, 57]. An intuitive idea is to transfer the trained lung area segmenter to the target data set, obtain the lung area, and use them as the constraints. However, according to domain knowledge, pneumothorax is an abnormal gas (lucent area) in the pleural space, which is localized in the lung + space on 2D chest radiographs. Additionally, the directly deployed lung segmenter suffers from a domain shift between the

source data set and the target data set [59, 60]. Consequently, certain portions of the generated lung area may exhibit detrimental noise to the downstream pneumothorax segmentation. To tackle the first challenge, we integrate several morphological operations into the fundamental lung area segmenter to derive the lung + space segmenter. To address the second issue, we design a lung + space discriminator to filter out the noisy constraints of lung + space and retain the informative ones.

2.2.2 | Phase 2: Lung + space segmentation

Based on the auxiliary lung segmenter, we implement three post-processing steps, including the largest connected component cutoff, morphological closing, and morphological dilation to develop a lung + space segmenter. The initial operation, the largest connected component cutoff, is employed to mitigate the impact of small islands, a common issue in medical image segmentation [61, 62]. By retaining the top two largest areas, we effectively extract the two frontal lung regions from the noisy segmentation output. Subsequently, morphological closing is applied to fill small holes inside the lung area, another prevalent concern in medical image segmentation [63, 64]. Lastly, morphological dilation is employed to extend the lung area, encompassing the side pleural space between the lung boundary and the chest wall [65, 66]. By combining these post-processing operations with the lung area segmenter, we have effectively crafted the lung + space segmenter which generates the candidate constraints of lung + space for each chest radiograph in the target data set of pneumothorax segmentation.

2.2.3 | Phase 3: Lung + space discrimination

Utilizing the lung + space segmenter, each chest radiograph I_i in the target task has been enriched with a constraint C_i Which holds an identical shape as I_i. In a constraint C_i, pixels located within the lung + space are assigned a value of 1, while those outside the lung + space are marked as 0. Nevertheless, due to the problem of domain shift [52, 67], some of the constraints still present noise issues. If all constraints are unquestioningly incorporated in model training, the improvements achieved by well-behaved constraints can potentially be negated by the detrimental constraints. To address this challenge, a lung + space discriminator is developed to differentiate whether a predicted lung + space contains significant noise.

The constraint discrimination constitutes a binary classification task. For training the discriminator,

Equation (1) derives binary labels by calculating the coverage rate R_i of pneumothorax annotation S_i in relation to the lung + space constraint C_i. If R_i exceeds τ, the binary label B_i is designated as 1; otherwise, it is set to 0.

$$R_i = \frac{|C_i \cap S_i|}{|S_i|} \tag{1}$$

Regarding the discriminator input, we followed a previous study [68] to fuse the original chest radiograph I_i, the constraint C_i, and the masked chest radiograph $I_i \times C_i$ on the channel level. The model output is configured to yield a single value, predicting the probability of including the constraint in the training phase. With the inputs, output targets, and the binary classifier, we proceed with the standard image classification using cross-entropy loss [65]. After the binary classification, certain chest radiographs are augmented with the constraints of lung + space. For those samples whose constraints are excluded, we supplement them with all-one matrixes to align the data format and nullify the constraint effect. Finally, we obtain the anatomical constraints of lung + space for the downstream training of the constrained segmenter. It is noteworthy that the classification is performed exclusively on the training and validation sets of the target datasets, ensuring that the test set is unseen and there is no information leakage.

2.3 | Constrained segmentation

We first present a standard formulation of image segmenter training without any constraints, which serves as the baseline for our study. Subsequently, we elaborate on the constrained training approach, which includes an additional penalty term aimed at satisfying the imposed constraints.

In a typical training stage of a single disease segmenter, we consider a data set D consisting of N input images I_i, $i = 1, ..., N$ and their respective lesion segmentations S_i, $i = 1, ..., N$. Then the data set D is randomly split into training set D_{train}, validation set D_{valid}, and test set D_{test} with N_{train}, N_{valid}, and N_{test} samples, respectively. After that, a model Y with parameter θ is trained on D_{train}, by minimizing the overall loss L averaging sample-wise loss l such as Dice or cross-entropy [69] between the model output $Y(I_i, \theta)$ and the ground-truth mask S_i.

To avoid the overfitting of Y, the optimization of θ is early-stopped upon reaching the loss plateau on D_{valid}. However, the loss function defined by Equation (2) overlooks the domain knowledge and trains the lesion segmenter through an end-to-end approach to map I_i to S_i. In clinical medicine, some diseases highly occur in certain

regions, which receive more attention than other areas in the diagnostic process [65, 66]. Therefore, the prior knowledge of the disease occurrence area potentially contributes to the training process of the disease segmenter.

$$L_{D_{\text{train}}} = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} l(Y(I_i, \theta), S_i). \tag{2}$$

2.3.1 | Phase 4: Constrained lesion area segmentation

In our proposed formulation, the disease occurrence area is introduced as a constraint in the loss function. To guide the model's focus on the disease occurrence area, the adjusted loss function defined by Equation (3) penalizes the model if the constraint is violated. Specifically, the loss function in Equation (2) will be supplemented with a novel penalty term P comparing the model output $Y(I_i, \theta)$ with a sample-specific constraint C_i:

$$\begin{aligned} L_{D_{\text{train}}} = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} (l(Y(I_i, \theta), S_i) \\ + \lambda \times P(Y(I_i, \theta), C_i)), \end{aligned} \tag{3}$$

where $l(Y(I_i, \theta), S_i)$ stands for the classic loss function for image segmentation in Equation (2), λ is a positive hyper-parameter fine-tuned on D_{valid}, and Equation (4) denotes the proposed penalty term $P(Y(I_i, \theta), C_i)$:

$$P(Y(I_i, \theta), C_i) = 1 - \frac{|Y(I_i, \theta) \cap C_i|}{|Y(I_i, \theta)|} \tag{4}$$

with $|Y(I_i, \theta) \cap C_i|$ standing for the intersection area size between $Y(I_i, \theta)$ and C_i, and $|Y(I_i, \theta)|$ representing the size of $Y(I_i, \theta)$. Clearly, when an outputted disease segmentation crosses the constraint boundary, the penalty function is positive while a satisfied segmentation within the constraint area corresponds to a null penalty. By adding such a differentiable term in the loss function, we address the difficulty of conventional convex-optimization of neural networks for the constraint satisfaction [70, 71].

2.4 | Experimental settings

2.4.1 | Datasets

Our proposed formulation was based on three external datasets of lung segmentation and a target data set of pneumothorax segmentation. Table 1 gives an overview

of the used datasets and their purposes in our experiments. The lung segmentation datasets were used to develop the auxiliary lung segmenter, the foundation for the lung + space segmenter. The pneumothorax data set was our target data set to build the lung + space discriminator and compare the pneumothorax segmenter trained by the baseline or the constrained loss function.

Lung segmentation data set

We developed the lung segmenter using JSRT, Shenzhen, and MC datasets. The JSRT data set was collected by the Japan Radiological Society and it includes 247 chest radiographs with a resolution of 2048×2048. The original data set contained no ground-truth lung area annotation, which was supplemented by another team [73]. The MC data set was gathered by Montgomery County's tuberculosis screening program and it comprises 138 chest radiographs with either a resolution of 4020×4892 or 4892×4020. The Shenzhen data set was built by Shenzhen No. 3 Hospital and it consists of 566 chest radiographs with varying resolutions around 3000×3000. Both MC and Shenzhen datasets were annotated with lung area masks by researchers from the U.S. National Institutes of Health [57]. We resized all chest radiographs from the three datasets and their corresponding lung area annotations into the resolution of 224×224 to comply with most pre-trained backbones [64, 74]. Then we employed the proportion of 70/20/10 for the split of training (173 JSRT chest radiographs, 396 Shenzhen chest radiographs, 97 MC chest radiographs), validation (49 JSRT chest radiographs, 113 Shenzhen chest radiographs, 28 MC chest radiographs), and test data set (25 JSRT chest radiographs, 57 Shenzhen chest radiographs, 13 MC chest radiographs). We included more chest radiographs in the validation data set than the test data set per the auxiliary task to develop a superior capstone for downstream lung + space segmentation.

Pneumothorax segmentation data set

With the auxiliary lung segmenter built on the external datasets, we trained the lung + space discriminator, the baseline pneumothorax segmenter, and the constrained pneumothorax segmenter on the Society for Imaging Informatics in Medicine (SIIM)-American College of Radiology (ACR) Pneumothorax Segmentation data set (SIIM-ACR) [72]. As a subset of the ChestX-ray14 data set [75], the SIIM-ACR data set was prepared by radiologists from SIIM and the Society of Thoracic Radiology. It comprises 2391 pneumothorax-positive chest radiographs and matching lesion annotations with a resolution of 1024×1024. All chest radiographs and lesion annotations were resized into 224×224 and further split into training (1674 chest radiographs), validation (239 chest radiographs), and test data set (478 chest radiographs) using the ratio of 70/10/20 to involve more samples in the test set and ensure the robustness of evaluation.

2.4.2 | Baseline segmentation

Motivated by the previous study demonstrating that classic architectures effectively disentangle the impacts of training strategies [76], we employed six segmentation networks using three well-established architectures of U-Net [38], LinkNet [77], or PSPNet [78], along with two backbones of VGG-11 [21] or MobileOne-S0 [74] to serve as the baseline. Due to fixed input formats that preclude additional constraints and limited computational resources, automated machine learning frameworks like nnU-Net [79] were not implemented. Furthermore, models that require human involvement, such as the segmentation anything model [80], were excluded because of the lack of necessary interaction data, specifically segmentation prompts. The model input was a grayscale chest radiograph, replicated three times on the channel level to align with most backbones. The model output was a standard single-channel probability map with a resolution of 224×224, which was binarized using a common threshold of 0.5 to delineate the specific disease region [81]. The standard Dice loss was used in model training [69].

TABLE 1 An overview of the used data set, abbreviation, and function.

Data set name	Abbreviation	Purpose
Japanese Society of Radiological Technology data set [56]	JSRT	Lung segmentation
Shenzhen data set [57]	Shenzhen	Lung + space segmentation
Montgomery County data set [57]	MC	
Society for Imaging Informatics in Medicine-American College of Radiology Pneumothorax Segmentation data set [72]	SIIM-ACR	Lung + space discrimination
		Pneumothorax segmentation

2.4.3 | Anatomical constraint learning

To learn well-behaved constraints of lung + space without additional annotation in the target data set, we followed a multi-step approach: First, we extracted constraints from an auxiliary task involving lung area segmentation using three external datasets; Next, we applied three morphological operations to transform the lung area into the lung + space. Finally, we trained a lung + space discriminator to filter out noisy constraints.

The lung segmenter in the first step was constructed based on the U-Net architecture with the VGG-11 backbone. The three morphological operations in the second step consisted of the following: the top two largest connected component cutoffs, closing with a 19×19 ellipse element [61] to fill voids within the lung area, and dilation with a 15×15 ellipse element [61] to smooth the lung area boundary and encompass the lung + space outside the lung area. The lung + space discriminator in the third step was based on the VGG-11 backbone, and we adjusted its output to produce a single predictive value. The coverage rate τ in Equation (1) was set as 0.99 to generate the classification label B_i. A further step for noise filtering was to increase the binarization cutoff value to include the constraints with high confidence. We chose the cutoff value based on several high specificity values of 0.80, 0.85, 0.90, and 0.95 and subsequently optimized it as a hyper-parameter in the constrained segmentation training.

2.4.4 | Constrained segmentation

After the selection via the lung + space discriminator, the well-behaved constraints were kept, and for those samples without constraints, all-one matrixes without penalty effect were used to ensure the data alignment in the training process. We modified the loss function in model training by adding a penalty term of Equation (4) to the standard Dice loss. The penalty term compared each pneumothorax prediction with its corresponding constraint. Hyper-parameter λ in the modified loss function of Equation (3) was grid selected from 0.2, 0.4, 0.6, 0.8, and 1.0 according to model performance on the validation set. Despite the modification of the loss function, all other settings remained unchanged as the baseline models, including the model inputs, outputs, architectures, backbones, and optimizer, to facilitate a fair comparison.

2.4.5 | Implementation details

We utilized the standard Dice loss to train the lung area segmenter and the baseline pneumothorax segmenter. For the training of the lung + space discriminator, we employed the classic cross-entropy loss. For the constrained training of the pneumothorax segmenter, we added a penalty term to the standard Dice loss. Across all experiments, stochastic gradient descent (SGD) was employed to minimize the respective loss functions. We initiated the learning rate at 0.01 and reduced it to 0.9 of its current value if no improvement was observed for five epochs on the validation data set. Table 2 provides an overview of the experimental settings.

For reproducibility, the pipeline was developed in PyTorch 1.12.1, and the code has been made open access [82]. We implemented the experiments on a Dell Precision 7920 Tower Workstation with an Intel Xeon Silver 4210 CPU and an NVIDIA GeForce RTX 2080 Super GPU.

2.4.6 | Evaluation metrics

Evaluation of constraints plausibility
The anatomical constraints of lung + space, generated by the external lung segmenter and morphological operations, contained high uncertainty. To filter out the noisy constraints, we trained a lung + space discriminator and

TABLE 2 Default experimental settings in constrained segmentation.

Phase	Hyper-parameter	Candidate
Auxiliary lung segmentation	Architecture	U-Net
	Backbone	VGG-11
	Loss function	Dice loss
	Optimizer	SGD
Lung + space segmentation	Closing element size	19×19
	Dilation element size	15×15
Lung + space discrimination	Cover rate	0.99
	Backbone	VGG-11
	Loss function	Cross-entropy
	Optimizer	SGD
Pneumothorax segmentation	Architecture	U-Net, LinkNet, PSPNet
	Backbone	VGG-11, MobileOne-S0
	Loss function	Dice loss
	Optimizer	SGD

used the area under the receiver operating characteristic curve (AUROC) to assess the model performance. We set the classification thresholds according to the specificity values of 0.80, 0.85, 0.90, and 0.95 [83] and reported the sensitivity, positive predictive value (PPV), and negative predictive value (NPV) to present the characteristics of predicted constraints after selection. Besides the evaluation metrics, we also reported their respective confidence intervals via bootstrapping.

Evaluation of segmentation performance
We utilized the intersection over union (IoU) [84], dice similarity coefficient (DSC) [85], and the Hausdorff distance (HD) [86] to assess the pixel-level difference between the predicted area and the ground-truth annotation in both the auxiliary task of lung area segmentation and the target task of pneumothorax segmentation. IoU and DSC are designed to quantify the degree of overlap between the predicted area and the ground-truth annotation. Accordingly, elevated values in these two metrics indicate enhanced model performance. On the other hand, HD assesses the distance between the two aforementioned regions, and therefore, diminished HD stands for better performance. In addition to the mean values on the test data set, we also provided the confidence intervals based on bootstrapping.

3 | RESULTS

This section presents both quantitative and qualitative results of each module in the proposed constrained segmentation. For the quantitative evaluations, we reported the results of auxiliary lung segmentation, lung + space discrimination, and pneumothorax segmentation. An ablation study was presented to underscore the efficacy of each designed element. We further highlighted the robustness study to validate the stability of our constraint-based formulation. For the qualitative assessments, we first showed the visual samples of lung segmentation derived from the external lung

segmentation data set. We then presented the various phases involved in lung + space generation. Lastly, we compared the pneumothorax segmenter trained by using baseline and that trained using our constrained loss function.

3.1 | Auxiliary lung area segmentation

Table 3 shows the auxiliary lung segmentation results using the U-Net architecture and the VGG-11 backbones. The segmenter yielded the best IoU, DSC, and HD on the test data set of the MC database and performed comparably well on the other two databases. Figure 2 visualizes three random samples and their segmentation results, demonstrating the outstanding performance of the auxiliary lung area segmenter on the external datasets.

3.2 | Lung + space discrimination

To obtain the lung + space, we first applied the auxiliary lung segmenter on the target data set of pneumothorax segmentation to obtain the initial lung area and then processed the lung area with three morphological operations. The first operation selected the top two largest connected components, thereby eliminating noisy segment islands [87]. The second operation entailed a closing operation using an element size of 19×19 to fill voids within the lung area. The final operation consisted of dilation with a 15×15 element size to encompass the pleural space situated between the lungs and chest wall.

Figure 3 presents different stages of lung segmentation and morphological operations on the SIIM-ACR data set. Chest radiographs in the first two rows demonstrate the noisy small islands and the holes in the lung structure, which is alleviated by the morphological operations. However, it is worth noting that not all constraints can be adjusted, as depicted in the last two rows, which showcase collapsed scenarios and validate the essential

TABLE 3 Performance of the lung segmenter on the external lung segmentation test datasets of JSRT, Shenzhen, and MC, in terms of IoU, DSC, and HD.

Datasets	IoU	DSC	HD
JSRT	0.949 (0.943–0.955)	0.974 (0.972–0.976)	3.814 (3.536–4.092)
Shenzhen	0.918 (0.902–0.934)	0.956 (0.946–0.966)	4.158 (3.807–4.509)
MC	0.961 (0.955–0.967)	0.980 (0.976–0.984)	3.732 (3.432–4.032)

Note: The mean value of each measurement is presented alongside its corresponding confidence interval, enclosed within brackets. Higher DSC or IoU values indicate better performance, while lower HD values demonstrate better performance.

Abbreviations: DSC, dice similarity coefficient; HD, Hausdorff distance; IoU, intersection over union.
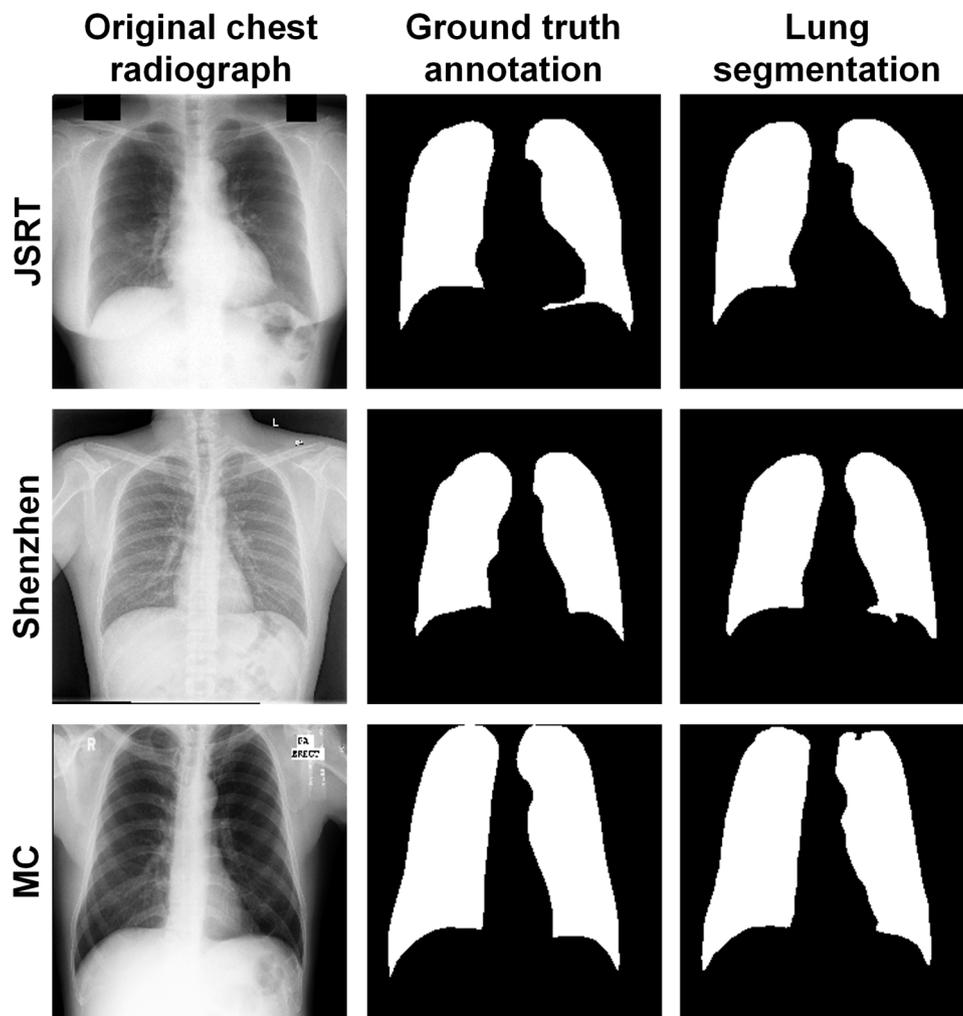
FIGURE 2 Visual samples of lung area segmentation in JSRT, Shenzhen, and MC databases.

role of the reliability discriminator in identifying and discarding erroneous constraints.

Based on the predicted lung + space, we compared it to the ground-truth pneumothorax annotation. A lung + space was labeled well-behaved if it covered 0.99 of its corresponding pneumothorax areas. With these binary labels, a discriminator was trained and tested. To further mitigate the introduction of noisy constraints in downstream training, we established cutoff values based on multiple specificity thresholds. Table 4 presents the discriminator's performance in terms of AUROC, specificity-based cutoff values, the respective specificity, sensitivity, PPV, and NPV values.

3.3 | Pneumothorax segmentation

Table 5 quantitatively compares the constrained and the baseline segmentation performance across different combinations of architectures and backbones. The constrained version consistently outperformed the baseline method and yielded average performance gains of 4.6%, 3.6%, and 3.3% in terms of IoU, DSC, and HD. Notably, our strategy achieved statistically significant improvements at the 0.05 level for HD on U-Net and PSPNet architectures with VGG-11 backbone, as evidenced by the nonoverlapping 95% confidence intervals shown in Table 5 [88]. While nonoverlapping confidence intervals were not observed in other scenarios, the constrained training strategy consistently improved mean values of IoU, DSC, and HD across all architectures and backbones. Moreover, U-Net architecture with VGG-11 backbone achieved the best performance in terms of IoU and DSC across all baseline models. Furthermore, U-Net architecture achieved better results compared with the relatively sophisticated LinkNet or PSPNet in most scenarios, demonstrating the effectiveness of U-Net for pneumothorax segmentation.

Figure 4 provides a comparative visualization of pneumothorax segmentation between the constrained
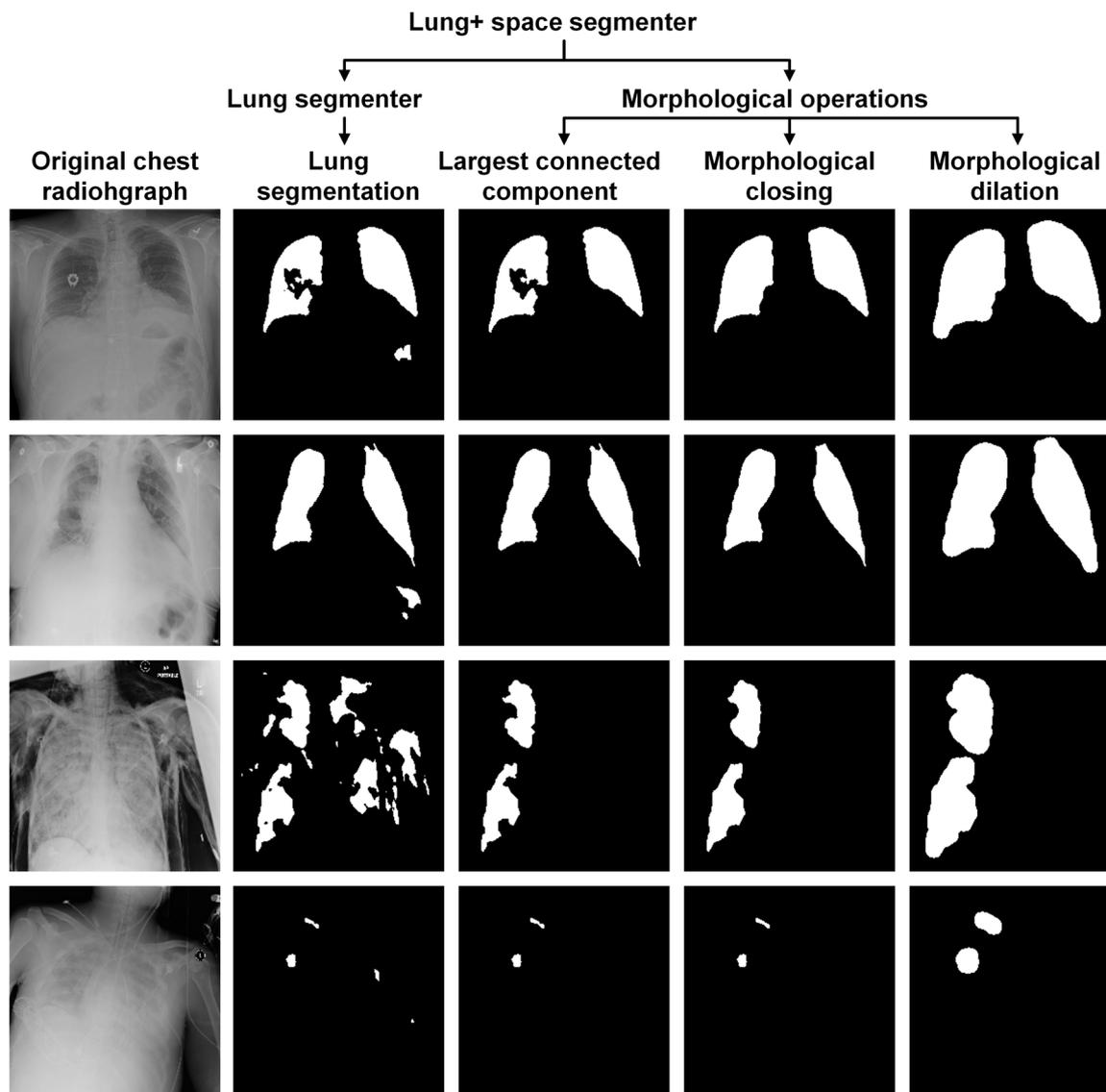
FIGURE 3 Visualization of different generation stages of the constraints of lung + space in SIIM-ACR database.

TABLE 4 Classification results of the lung + space segmentation discriminator on the test data set of the SIIM-ACR data set.

Cutoff value[a]	AUROC	Specificity	Sensitivity	PPV	NPV
0.68	0.750 (0.705–0.795)	0.854 (0.793–0.915)	0.394 (0.333–0.455)	0.846 (0.787–0.905)	0.410 (0.353–0.467)
0.70		0.924 (0.875–0.973)	0.244 (0.193–0.295)	0.867 (0.787–0.947)	0.376 (0.323–0.429)
0.70		0.924 (0.875–0.973)	0.244 (0.193–0.295)	0.867 (0.787–0.947)	0.376 (0.323–0.429)
0.72		0.962 (0.935–0.989)	0.147 (0.114–0.180)	0.887 (0.813–0.961)	0.358 (0.309–0.407)

Note: The mean value of each metric is presented alongside its corresponding confidence interval, enclosed within brackets.

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

[a]Cutoff value was incrementally determined with a step of 0.1, utilizing specificity thresholds of 0.80, 0.85, 0.90, and 0.95 on the validation set. In our experiments, as cutoff values were escalated, shifts in specificity values were observed on the validation set, resulting in identical cutoff values across distinct specificity thresholds of 0.85 and 0.90.

and baseline segmenters using the architecture of U-Net and the backbone of VGG-11. The first row exemplifies a scenario where the proposed method surpasses the baseline model, the second row presents a scenario where the proposed method achieves performance on par with the baseline model, and the third row outlines a situation where the proposed method underperformed the baseline model. The last

TABLE 5 Performance comparison of the constrained and baseline segmentation with different architectures and backbones, in terms of IoU, DSC, and HD.

Architectures	Backbones	Methods	IoU	DSC	HD
U-Net	VGG-11	Baseline	0.316 (0.296–0.336)	0.441 (0.417–0.465)	4.799 (4.681–4.917)
		Ours	0.336 (0.316–0.356)	0.461 (0.437–0.485)	4.558 (4.454–4.662)
		Improvement	6.3%	4.5%	5.0%
	MobileOne-S0	Baseline	0.309 (0.287–0.331)	0.431 (0.404–0.458)	4.703 (4.605–4.801)
		Ours	0.326 (0.306–0.346)	0.449 (0.427–0.471)	4.586 (4.496–4.676)
		Improvement	5.5%	4.2%	2.5%
LinkNet	VGG-11	Baseline	0.305 (0.287–0.323)	0.426 (0.404–0.448)	4.740 (4.652–4.828)
		Ours	0.322 (0.300–0.344)	0.447 (0.422–0.472)	4.592 (4.490–4.694)
		Improvement	5.6%	4.9%	3.1%
	MobileOne-S0	Baseline	0.302 (0.284–0.320)	0.425 (0.403–0.447)	4.839 (4.743–4.935)
		Ours	0.320 (0.300–0.340)	0.447 (0.423–0.471)	4.675 (4.589–4.761)
		Improvement	6.0%	5.2%	3.4%
PSPNet	VGG-11	Baseline	0.302 (0.282–0.322)	0.424 (0.400–0.448)	4.866 (4.768–4.964)
		Ours	0.307 (0.289–0.325)	0.429 (0.407–0.451)	4.660 (4.558–4.762)
		Improvement	1.7%	1.2%	4.2%
	MobileOne-S0	Baseline	0.260 (0.242–0.278)	0.377 (0.355–0.399)	5.008 (4.900–5.116)
		Ours	0.267 (0.247–0.287)	0.382 (0.358–0.406)	4.935 (4.831–5.039)
		Improvement	2.7%	1.3%	1.5%
Average improvement (%)			4.6	3.6	3.3

Note: The mean value of each measurement is presented alongside its corresponding confidence interval, enclosed within brackets. Higher DSC or IoU values indicate better performance, while lower HD values demonstrate better performance.

Abbreviations: DSC, dice similarity coefficient; HD, Hausdorff distance; IoU, intersection over union.

row illustrates cases where both methods encounter difficulties.

3.4 | Ablation study

To ascertain the efficacy of each component within the proposed formulation, we conducted an ablation study on the U-Net architecture and the VGG-11 backbone. As shown in Table 6, the constraints generated by the external lung area segmenter exhibited excessive noise, resulting in subpar performance compared to the baseline without any constraints. While training with constraints derived from the lung + space segmenter yielded better performance, it only achieved performance marginally better than the baseline. In contrast, with the integration of a lung + space discriminator, the noisy constraints were efficiently filtered, preserving the well-behaved ones, and leading to substantial improvements across all evaluation metrics.

3.5 | Robustness to the constraint settings

Finally, we investigated the robustness of hyper-parameters in anatomical constraint learning. To this end, we validated a range of values in the morphological sizes and cover rates. In each validation, the architecture was configured as U-Net, employing VGG-11 as the specific backbone, while all other parameters were maintained at their default values, as outlined in Table 2. Tables 7 and 8 detail the ablation results of various morphological sizes and cover rates. As anticipated, the constrained model consistently exhibited superior performance than the baseline model across varying closing element sizes, dilation element sizes, and coverage rates. Notably, the closing element size of 25×25 and the dilation element size of 20×20 in Table 7 attained a new state-of-the-art performance, surpassing the default parameter settings [61]. Similarly, the cover rate of 0.90 in Table 8 presented better performance than the default setting of 0.99. These observations suggest that our proposed

Original chest radiograph	Ground truth annotation	Constrained segmentation	Baseline segmentation
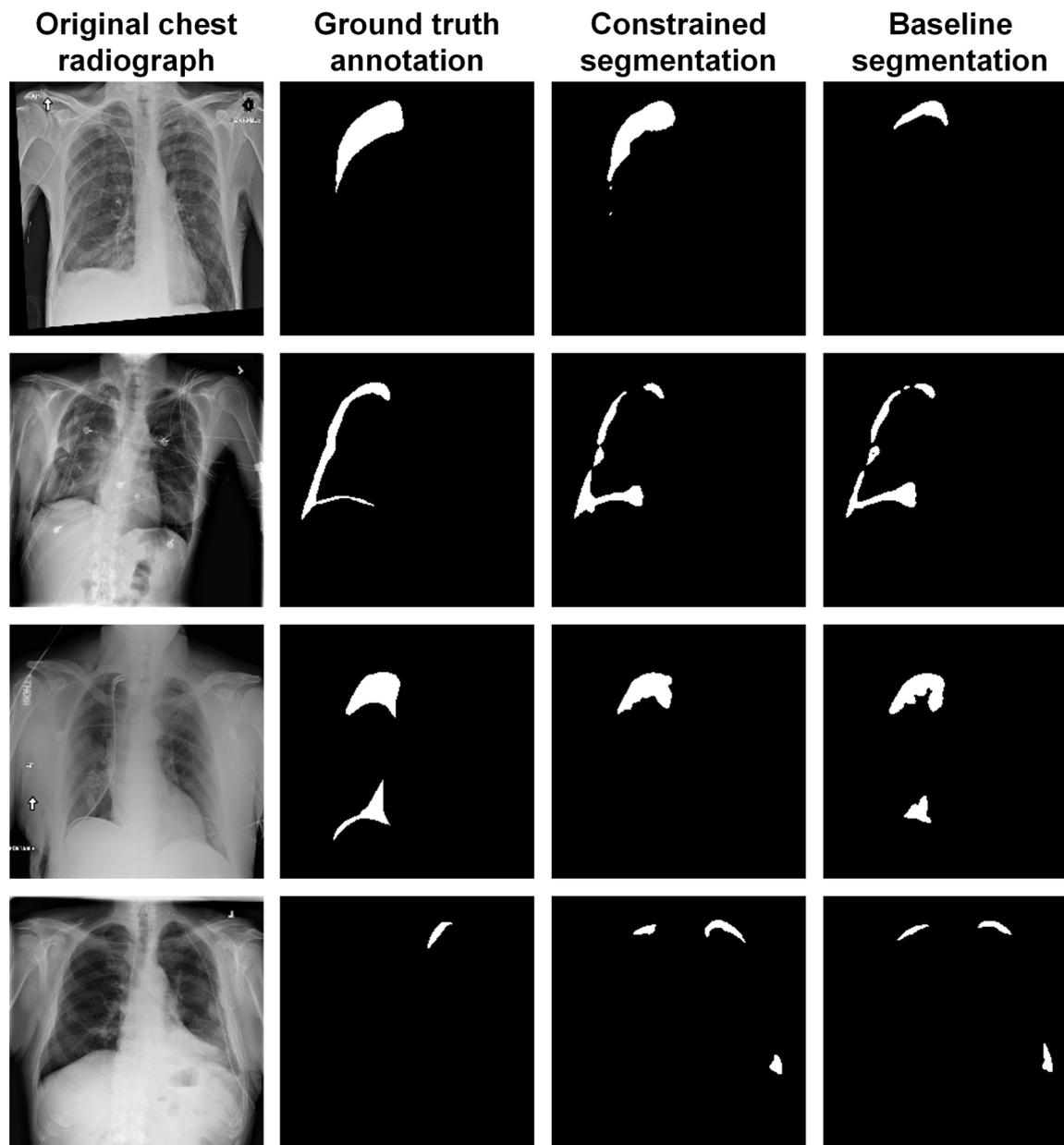
FIGURE 4 Comparative example of the constrained segmentation and the baseline segmentation. Rows 1–4 represent the following scenarios: the constrained method outperforming the baseline, the constrained method achieving comparable results to the baseline, the constrained method underperforming the baseline, and both methods collapsing, respectively.

TABLE 6 Ablation study on the lung + space segmentation model and the lung + space segmentation discriminator.

Methods	Lung area segmenter	Lung + space segmenter	Lung + space discriminator	IoU	DSC	HD
Baseline	×	×	×	0.316 (0.296–0.336)	0.441 (0.417–0.465)	4.799 (4.681–4.917)
	√	×	×	0.298 (0.278–0.318)	0.421 (0.397–0.445)	4.895 (4.791–4.999)
	×	√	×	0.317 (0.297–0.337)	0.439 (0.415–0.463)	4.770 (4.682–4.858)
Ours	×	√	√	0.336 (0.316–0.356)	0.461 (0.437–0.485)	4.558 (4.454–4.662)

Abbreviations: DSC, dice similarity coefficient; HD, Hausdorff distance; IoU, intersection over union.

TABLE 7 Robustness study on morphological element sizes in the lung + space segmenter.

Methods	Closing element size	Dilation element size	IoU	DSC	HD
Baseline	×	×	0.316 (0.296–0.336)	0.441 (0.417–0.465)	4.799 (4.681–4.917)
Ours	15 × 15	10 × 10	0.331 (0.311–0.351)	0.458 (0.434–0.482)	4.691 (4.603–4.779)
		15 × 15	0.338 (0.316–0.360)	0.462 (0.438–0.486)	4.439 (4.343–4.535)
		20 × 20	0.338 (0.318–0.358)	0.463 (0.439–0.487)	4.538 (4.438–4.638)
	19 × 19[a]	10 × 10	0.332 (0.312–0.352)	0.455 (0.431–0.479)	4.521 (4.435–4.607)
		15 × 15[a]	0.336 (0.316–0.356)	0.461 (0.437–0.485)	4.558 (4.454–4.662)
		20 × 20	0.340 (0.320–0.360)	0.465 (0.441–0.489)	4.488 (4.374–4.602)
	25 × 25	10 × 10	0.339 (0.317–0.361)	0.463 (0.436–0.490)	4.442 (4.364–4.520)
		15 × 15	0.337 (0.319–0.355)	0.465 (0.441–0.489)	4.570 (4.482–4.658)
		20 × 20	0.346 (0.324–0.368)	0.473 (0.448–0.498)	4.372 (4.280–4.464)

Abbreviations: DSC, dice similarity coefficient; HD, Hausdorff distance; IoU, intersection over union.

[a]Parameter settings in the prior literature [61].

TABLE 8 Robustness study on cover rates in the lung + space discriminator.

Methods	Cover rate	IoU	DSC	HD
Baseline	×	0.316 (0.296–0.336)	0.441 (0.417–0.465)	4.799 (4.681–4.917)
Ours	0.80	0.334 (0.314–0.354)	0.460 (0.436–0.484)	4.408 (4.306–4.510)
	0.90	0.343 (0.321–0.365)	0.470 (0.445–0.495)	4.474 (4.380–4.568)
	0.99	0.336 (0.316–0.356)	0.461 (0.437–0.485)	4.558 (4.454–4.662)

Abbreviations: DSC, dice similarity coefficient; HD, Hausdorff distance; IoU, intersection over union.

constrained training could be further enhanced through hyper-parameter optimization.

4 | DISCUSSION

Conventional DL-based lesion segmentation models typically employ an end-to-end approach, directly mapping input medical images into lesion delineations without accounting for clinical knowledge, such as the spatial distribution of the disease. In this study, we introduced a loss function that incorporates disease occurrence area as a constraint into the standard training framework. Through numerical studies, we demonstrated the effectiveness of the proposed training strategy on pneumothorax segmentation, a condition sensitive to location, manifesting in the pleural space between the lungs and the chest wall. Consistently, our proposed training method outperformed the baseline approach across three different architectures paired with two encoders. Moreover, our proposed method can be seamlessly integrated with the existing segmentation studies focusing on model architectures or training strategies.

For instance, Wang et al. developed an ensemble approach integrating two architectures with four encoders based on Dice loss [40]. Abedalla et al. introduced a two-stage method for pneumothorax segmentation: first training a U-Net on low-resolution chest radiographs, then fine-tuning it with high-resolution images using a combination of Dice and cross-entropy loss [39]. These solutions may benefit from adding our proposed penalty term to their original loss functions [89].

While our method was showcased for pneumothorax segmentation, its applicability extends to other thoracic diseases with location-relevant characteristics [90]. Many of these conditions exhibit location-sensitive characteristics within the lung area [91, 92]. For thoracopathy localized in the lung area, removing the morphological dilation and utilizing the resulting lung area can potentially improve the model performance using our proposed constrained training strategy. Additionally, the proposed method can be adapted to thoracopathy localization tasks that require a bounding box around the lesion area pre-specified by experts. A two-step method would be to first generate pixel-level constraints of lung area and identify the rectangular boxes encompassing

these constraints, and then to integrate the penalty term in Equation (3) into the classic $L2$ distance-based loss function, ensuring the model's output remains within the bounding box-based constraints.

In contrast to the previous constraint learning methods [53, 67] that required additional annotation on the target data set and task, our method employs a four-phase pipeline to generate sample-specific constraints. This sets a valuable benchmark for the development of constrained models in various domains. Our method can be generalized by (1) generating initial constraints based on an auxiliary task and external datasets, (2) refining these initial constraints based on the relationship between the auxiliary and the target tasks, and (3) employing the target data set and available annotations to filter out noisy constraints, retaining only the informative ones. As highlighted in Table 6, our ablation experiment indicates that the initial constraints transferred from external datasets can introduce detrimental noise, potentially hindering model convergence. However, after refining and selecting constraints, the model's performance improves. Tables 7 and 8 further demonstrate the robustness of our method to hyper-parameter variations in the second and third stages of learning constraints. In fact, segmenters trained under various settings outperformed the ones under the default setting, suggesting a promising avenue for future hyper-parameter optimization.

While the proposed method has demonstrated consistent and robust improvements, there are several limitations to be addressed in future work. First, our model involves multiple hyper-parameters. Although we have made initial adjustments, there is potential for further optimization. Intuitively, we hypothesize that there exists a certain relationship between the hyper-parameters [79]. For instance, extensive morphological operations can broaden constraints' boundaries, suggesting a need to adjust the cover rate to mitigate noise from these expanded boundaries [93]. Second, while this work primarily incorporates anatomical shapes as constraints, future research will delve into other geometric attributes such as sphericity, convexity, and roundness [94, 95]. It is of interest to explore their impact on pneumothorax segmentation. Third, the present study was restricted by the limited computational resources, the lack of segmentation prompts, the restricted access to commercial software, the absence of imaging conditions and diverse modalities, and the vacancy of patient demographics, hindering the implementation of automated machine learning frameworks such as nnU-Net [79], interactive models such as segmentation anything model [80], commercial segmentation solutions such as Siemens AI-Rad Companion [96, 97], robustness studies on imaging conditions and diverse modalities, and subpopulation evaluations [98]. Future research will collaborate with ML engineers and medical experts to launch these experiments and explore how trained physicians can benefit from these DL tools [44, 99]. Lastly, previous studies demonstrate a strong correlation between anatomical information and the diagnosis of various thoracic diseases [62, 100]. Therefore, we plan to evaluate the adaptability of the proposed method across a broader spectrum of thoracopathy tasks, including classification, detection, and segmentation [101, 102].

5 | CONCLUSIONS

Historically, domain knowledge was underemphasized by the DL community when tackling medical tasks. In this study, we underscore the value of integrating clinical knowledge, particularly regarding disease occurrence, to enhance DL-based pneumothorax segmentation. Different from previous work that requires additional annotation on the target data set and task, our approach leverages external open-access datasets and an auxiliary task. This strategy not only streamlines our process but also offers versatility, making it a promising framework for diagnosing other thoracic conditions or diseases with location-relevant characteristics.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no data sets were generated or analyzed during the current study.

ETHICS STATEMENT

Ethics approval was not required for this study as it utilized retrospective datasets that are publicly accessible. Researchers seeking access to the original data should request permission from the data owners and comply with their established protocols on data privacy and confidentiality.

INFORMED CONSENT

Not applicable.

ORCID

Han Yuan http://orcid.org/0000-0002-2674-6068
Nan Liu http://orcid.org/0000-0003-3610-4883

REFERENCES

1. Imran JB, Eastman AL. Pneumothorax. JAMA. 2017;318(10): 974. https://doi.org/10.1001/jama.2017.10476
2. Sahn SA, Heffner JE. Spontaneous pneumothorax. N Engl J Med. 2000;342(12):868–74. https://doi.org/10.1056/NEJM200003233421207
3. Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: a survey. Med Image Anal. 2021;72:102125. https://doi.org/10.1016/j.media.2021.102125
4. Noppen M, De Keukeleire T. Pneumothorax. Respiration. 2008;76(2):121–7. https://doi.org/10.1159/000135932
5. Ding W, Shen Y, Yang J, He X, Zhang M. Diagnosis of pneumothorax by radiography and ultrasonography. Chest. 2011;140(4):859–66. https://doi.org/10.1378/chest.10-2946
6. Gu D, Su K, Zhao H. A case-based ensemble learning system for explainable breast cancer recurrence prediction. Artif Intell Med. 2020;107:101858. https://doi.org/10.1016/j.artmed.2020.101858
7. Hong W, Hwang EJ, Lee JH, Park J, Goo JM, Park CM. Deep learning for detecting pneumothorax on chest radiographs after needle biopsy: clinical implementation. Radiology. 2022;303(2):433–41. https://doi.org/10.1148/radiol.211706
8. Sharma N, Aggarwal L. Automated medical image segmentation techniques. J Med Phys. 2010;35(1):3–14. https://doi.org/10.4103/0971-6203.58777
9. Duncan JS, Ayache N. Medical image analysis: progress over two decades and the challenges ahead. IEEE Trans Pattern Anal Mach Intell. 2000;22(1):85–106. https://doi.org/10.1109/34.824822
10. McInerney T, Terzopoulos D. Deformable models in medical image analysis: a survey. Med Image Anal. 1996;1(2):91–108. https://doi.org/10.1016/s1361-8415(96)80007-7
11. Pham DL, Xu C, Prince JL. Current methods in medical image segmentation. Annu Rev Biomed Eng. 2000;2:315–37. https://doi.org/10.1146/annurev.bioeng.2.1.315
12. Grau V, Mewes AUJ, Alcaniz M, Kikinis R, Warfield SK. Improved watershed transform for medical image segmentation using prior information. IEEE Trans Med Imaging. 2004;23(4):447–58. https://doi.org/10.1109/TMI.2004.824224
13. Ahmed MN, Yamany SM, Mohamed N, Farag AA, Moriarty T. A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. IEEE Trans Med Imaging. 2002;21(3):193–9. https://doi.org/10.1109/42.996338
14. Xie F, Yuan H, Ning Y, Ong MEH, Feng M, Hsu W, et al. Deep learning for temporal data representation in electronic health records: a systematic review of challenges and methodologies. J Biomed Inf. 2022;126:103980. https://doi.org/10.1016/j.jbi.2021.103980
15. Wang Q, Liu Q, Luo G, Liu Z, Huang J, Zhou Y, et al. Automated segmentation and diagnosis of pneumothorax on chest X-rays with fully convolutional multi-scale ScSE-DenseNet: a retrospective study. BMC Med Inform Decis Mak. 2020;20(Suppl 14): 317. https://doi.org/10.1186/s12911-020-01325-5
16. Iqbal T, Shaukat A, Akram MU, Mustansar Z, Khan A. Automatic diagnosis of pneumothorax from chest radiographs: a systematic literature review. IEEE Access. 2021;9: 145817–39. https://doi.org/10.1109/ACCESS.2021.3122998
17. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, et al. Deep learning in radiology. Academic Radiol. 2018;25(11):1472–80. https://doi.org/10.1016/j.acra.2018.02.018
18. Moses DA. Deep learning applied to automatic disease detection using chest X-rays. J Med Imaging Radiat Oncol. 2021;65(5):498–517. https://doi.org/10.1111/1754-9485.13273
19. Tang YX, Tang YB, Peng Y, Yan K, Bagheri M, Redd BA, et al. Automated abnormality classification of chest radiographs using deep convolutional neural networks. npj Digital Med. 2020;3:70. https://doi.org/10.1038/s41746-020-0273-z
20. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90. https://doi.org/10.1145/3065386
21. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd Int Conf Learn Represent ICLR 2015 Conf Track Proc. 2015:1–14. https://arxiv.org/abs/1409.1556
22. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015:1–9. https://doi.org/10.1109/CVPR.2015.7298594
23. Thian YL, Ng D, Hallinan JTPD, Jagmohan P, Sia SY, Tan CH, et al. Deep learning systems for pneumothorax detection on chest radiographs: a multicenter external validation study. Radiol: Artif Intell. 2021;3(4):e200190. https://doi.org/10.1148/ryai.2021200190
24. Tan M, Le QV EfficientNet: rethinking model scaling for convolutional neural networks. 2019. *arXiv.* http://arxiv.org/abs/1905.11946
25. 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 Bioinf. 2021;18(3):951–62. https://doi.org/10.1109/TCBB.2019.2911947
26. Feng S, Liu Q, Patel A, Bazai SU, Jin CK, Kim JS, et al. Automated pneumothorax triaging in chest X-rays in the New Zealand population using deep-learning algorithms. J Med Imaging Radiat Oncol. 2022;66(8):1035–43. https://doi.org/10.1111/1754-9485.13393
27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016:770–8. https://doi.org/10.1109/CVPR.2016.90

28. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017:2261–9. https://doi.org/10.1109/CVPR.2017.243

29. Mosquera C, Diaz FN, Binder F, Rabellino JM, Benitez SE, Beresñak AD, et al. Chest X-ray automated triage: a semiologic approach designed for clinical implementation, exploiting different types of labels through a combination of four deep learning architectures. Comput Methods Programs Biomed. 2021;206:106130. https://doi.org/10.1016/j.cmpb.2021.106130

30. Shvets AA, Rakhlin A, Kalinin AA, Iglovikov VI. Automatic Instrument Segmentation in Robot-Assisted Surgery using Deep Learning. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 2018:624–8. https://doi.org/10.1109/ICMLA.2018.00100

31. Rocha J, Pereira SC, Pedrosa J, Campilho A, Mendonça AM. Attention-driven spatial transformer network for abnormality detection in chest X-ray images. 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS), Shenzhen, China, 2022:252–7. https://doi.org/10.1109/CBMS55023.2022.00051

32. Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks. Montreal, Canada: Advances in Neural Information Processing Systems (NeurIPS); 2015. https://arxiv.org/abs/1506.02025

33. Chen KC, Yu HR, Chen WS, Lin WC, Lee YC, Chen HH, et al. Diagnosis of common pulmonary diseases in children by X-ray images and deep learning. Sci Rep. 2020;10(1):17374. https://doi.org/10.1038/s41598-020-73831-5

34. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016:779–88. https://doi.org/10.1109/CVPR.2016.91

35. Li Y, Wong D, Sreng S, Chung J, Toh A, Yuan H, et al. Effect of childhood atropine treatment on adult choroidal thickness using sequential deep learning-enabled segmentation. Asia-Pacific J Ophthalmol. 2024;13(5):100107. https://doi.org/10.1016/j.apjo.2024.100107

36. Saporta A, Gui X, Agrawal A, Pareek A, Truong SQH, Nguyen CDT, et al. Benchmarking saliency methods for chest X-ray interpretation. Nat Mach Intell. 2022;4:867–78. https://doi.org/10.1038/s42256-022-00536-x

37. Agrawal T, Choudhary P. Segmentation and classification on chest radiography: a systematic survey. Visual Comput. 2023;39(3):875–913. https://doi.org/10.1007/s00371-021-02352-7

38. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Lecture notes in computer science. Cham: Springer International Publishing; 2015. p. 234–41. https://doi.org/10.1007/978-3-319-24574-4_28

39. Abedalla A, Abdullah M, Al-Ayyoub M, Benkhelifa E. 2ST-UNet: 2-Stage Training Model using U-Net for pneumothorax segmentation in chest X-rays. 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020:1–6. https://doi.org/10.1109/IJCNN48605.2020.9207268

40. Wang X, Yang S, Lan J, Fang Y, He J, Wang M, et al. Automatic segmentation of pneumothorax in chest radiographs based on a two-stage deep learning method. IEEE Trans Cogn Dev Syst. 2022;14(1):205–18. https://doi.org/10.1109/TCDS.2020.3035572

41. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018;40(4):834–48. https://doi.org/10.1109/TPAMI.2017.2699184

42. He K, Gkioxari G, Dollár P, Girshick R, Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017:2980–8.

43. 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. 2021;25(5):1660–72. https://doi.org/10.1109/JBHI.2020.3023476

44. Lee SY, Ha S, Jeon MG, Li H, Choi H, Kim HP, et al. Localization-adjusted diagnostic performance and assistance effect of a computer-aided detection system for pneumothorax and consolidation. npj Digital Med. 2022;5(1):107. https://doi.org/10.1038/s41746-022-00658-x

45. Yuan H, Yu K, Xie F, Liu M, Sun S. Automated machine learning with interpretation: a systematic review of methodologies and applications in healthcare. Med Adv. 2024;2(3):205–37. https://doi.org/10.1002/med4.75

46. Pereira S, Meier R, McKinley R, Wiest R, Alves V, Silva CA, et al. Enhancing interpretability of automatically extracted machine learning features: application to a RBM-Random Forest system on brain lesion segmentation. Med Image Anal. 2018;44:228–44. https://doi.org/10.1016/j.media.2017.12.009

47. Yuan H, Kang L, Li Y, Fan Z. Human-in-the-loop machine learning for healthcare: current progress and future opportunities in electronic health records. Med Adv. 2024;2(3):318–22. https://doi.org/10.1002/med4.70

48. Van Ginneken B, Ter Haar Romeny BM, Viergever MA. Computer-aided diagnosis in chest radiography: a survey. IEEE Trans Med Imaging. 2001;20(12):1228–41. https://doi.org/10.1109/42.974918

49. Li F, Armato SG, Engelmann R, Rhines T, Crosby J, Lan L, et al. Anatomic point-based lung region with zone identification for radiologist annotation and machine learning for chest radiographs. J Digit Imaging. 2021;34(4):922–31. https://doi.org/10.1007/s10278-021-00494-7

50. Crosby J, Rhines T, Li F, MacMahon H, Giger M. Deep learning for pneumothorax detection and localization using networks fine-tuned with multiple institutional datasetsMedical Imaging 2020: Computer-Aided Diagnosis, Houston, USA, 2020. https://doi.org/10.1117/12.2549709

51. Jung HG, Nam WJ, Kim HW, Lee SW. Weakly supervised thoracic disease localization via disease masks. Neurocomputing. 2023;517:34–43. https://doi.org/10.1016/j.neucom.2022.10.019

52. Bateson M, Dolz J, Kervadec H, Lombaert H, Ayed IB. Constrained domain adaptation for image segmentation.

IEEE Trans Med Imaging. 2021;40(7):1875–87. https://doi.org/10.1109/TMI.2021.3067688

53. Bateson M, Kervadec H, Dolz J, Lombaert H, Ben Ayed I. Constrained domain adaptation for segmentation. 2019 International conference on medical image computing and computer-assisted intervention (MICCAI), Shenzhen, China, 2019:326-334. https://doi.org/10.1007/978-3-030-32245-8_37

54. Kervadec H, Dolz J, Tang M, Granger E, Boykov Y, Ben Ayed I. Constrained-CNN losses for weakly supervised segmentation. Med Image Anal. 2019;54:88–99. https://doi.org/10.1016/j.media.2019.02.009

55. van Ginneken B, Katsuragawa S, ter Haar Romeny BM, Kunio Doi D, Viergever MA. Automatic detection of abnormalities in chest radiographs using local texture analysis. IEEE Trans Med Imaging. 2002;21(2):139–49. https://doi.org/10.1109/42.993132

56. Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu K, et al. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. AJR Am J Roentgenol. 2000;174(1):71–4. https://doi.org/10.2214/ajr.174.1.1740071

57. Jaeger S, Candemir S, Antani S, Wáng YX, Lu PX, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg. 2014;4(6):475–7. https://doi.org/10.3978/j.issn.2223-4292.2014.11.20

58. Bateson M, Kervadec H, Dolz J, Lombaert H, Ben Ayed I. Source-relaxed domain adaptation for image segmentation. In: Lecture notes in computer science. Cham: Springer International Publishing; 2020. p. 490–9. https://doi.org/10.1007/978-3-030-59710-8_48

59. Huang X, Yang X, Dou H, Huang Y, Zhang L, Liu Z, et al. Test-time bi-directional adaptation between image and model for robust segmentation. Comput Methods Programs Biomed. 2023;233:107477. https://doi.org/10.1016/j.cmpb.2023.107477

60. Larrazabal AJ, Martinez C, Glocker B, Ferrante E. Post-DAE: anatomically plausible segmentation via post-processing with denoising autoencoders. IEEE Trans Med Imaging. 2020;39(12):3813–20. https://doi.org/10.1109/TMI.2020.3005297

61. Souza JC, Bandeira Diniz JO, Ferreira JL, França da Silva GL, Corrêa Silva A, de Paiva AC. An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. Comput Methods Programs Biomed. 2019;177:285–96. https://doi.org/10.1016/j.cmpb.2019.06.005

62. Candemir S, Jaeger S, Palaniappan K, Musco JP, Singh RK, Zhiyun Xue X, et al. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Trans Med Imaging. 2014;33(2):577–90. https://doi.org/10.1109/TMI.2013.2290491

63. Mansoor A, Bagci U, Xu Z, Foster B, Olivier KN, Elinoff JM, et al. A generic approach to pathological lung segmentation. IEEE Trans Med Imaging. 2014;33(12):2293–310. https://doi.org/10.1109/TMI.2014.2337057

64. Mansoor A, Cerrolaza JJ, Perez G, Biggs E, Okada K, Nino G, et al. A generic approach to lung field segmentation from chest radiographs using deep space and shape learning. IEEE Trans Biomed Eng. 2020;67(4):1206–20. https://doi.org/10.1109/TBME.2019.2933508

65. Yuan H, Jiang P, Zhao G. Human-guided design to explain deep learning-based pneumothorax classifier. 2023 International Conference on Medical Imaging with Deep Learning (MIDL), Nashville, TN, USA, 2023:1–4. https://openreview.net/forum?id=_kk8KI8MiRE

66. Yuan H, Hong C, Jiang PT, Zhao G, Tran NTA, Xu X, et al. Clinical domain knowledge-derived template improves post hoc AI explanations in pneumothorax classification. J Biomed Inf. 2024;156:104673. https://doi.org/10.1016/j.jbi.2024.104673

67. Bateson M, Kervadec H, Dolz J, Lombaert H, Ben Ayed I. Source-free domain adaptation for image segmentation. Med Image Anal. 2022;82:102617. https://doi.org/10.1016/j.media.2022.102617

68. Frank O, Schipper N, Vaturi M, Soldati G, Smargiassi A, Inchingolo R, et al. Integrating domain knowledge into deep networks for lung ultrasound with applications to COVID-19. IEEE Trans Med Imaging. 2022;41(3):571–81. https://doi.org/10.1109/TMI.2021.3117246

69. Ma J, Chen J, Ng M, Huang R, Li Y, Li C, et al. Loss odyssey in medical image segmentation. Med Image Anal. 2021;71:102035. https://doi.org/10.1016/j.media.2021.102035

70. Klodt M, Cremers D. A convex framework for image segmentation with moment constraints. 2011 International Conference on Computer Vision, Barcelona, Spain, 2011:2236–2243. https://doi.org/10.1109/ICCV.2011.6126502

71. Wang L, Wang L, Chen KC, Shi F, Liao S, Li G, et al. Automated segmentation of CBCT image using spiral CT atlases and convex optimization. Med Image Comput Comput Assist Interv. 2013;16(pt 3):251–8. https://doi.org/10.1007/978-3-642-40760-4_32

72. Langer SG, Shih G. SIIM-ACR Pneumothorax Segmentation Dataset. 2019. https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation

73. van Ginneken B, Stegmann MB, Loog M. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Med Image Anal. 2006;10(1):19–40. https://doi.org/10.1016/j.media.2005.02.002

74. Vasu PKA, Gabriel J, Zhu J, Tuzel O, Ranjan A. MobileOne: an improved one millisecond mobile backbone. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023:7907–7917. https://doi.org/10.1109/CVPR52729.2023.00764

75. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-Ray8:hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common Thorax diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017:3462–71. https://doi.org/10.1109/CVPR.2017.369

76. Gut D, Tabor Z, Szymkowski M, Rozynek M, Kucybala I, Wojciechowski W. Benchmarking of deep architectures for segmentation of medical images. IEEE Trans Med Imaging. 2022;41(11):3231–41. https://doi.org/10.1109/TMI.2022.3180435

77. Chaurasia A, Culurciello E. LinkNet: exploiting encoder representations for efficient semantic segmentation. 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 2017:1–4. https://doi.org/10.1109/VCIP.2017.8305148

78. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017:6230–6239. https://doi.org/10.1109/CVPR.2017.660

79. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203–11. https://doi.org/10.1038/s41592-020-01008-z

80. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment anything. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2023:3992–4003. https://doi.org/10.1109/ICCV51070.2023.00371

81. Cai Z, Vasconcelos N. Cascade R-CNN: high quality object detection and instance segmentation. IEEE Trans Pattern Anal Mach Intell. 2021;43(5):1483–98. https://doi.org/10.1109/TPAMI.2019.2956516

82. Yuan H. Constrained learning code. 2023. https://github.com/Han-Yuan-Med/constrained-learning

83. Foo LL, Lim GYS, Lanca C, Wong CW, Hoang QV, Zhang XJ, et al. Deep learning system to predict the 5-year risk of high myopia using fundus imaging in children. npj Digital Med. 2023;6(1):10. https://doi.org/10.1038/s41746-023-00752-8

84. Jaccard P. The distribution of the flora in the alpine zone.1. New Phytol. 1912;11:37–50. https://doi.org/10.1111/J.1469-8137.1912.TB05611.X

85. Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945;26(3):297–302. https://doi.org/10.2307/1932409

86. Beauchemin M, Thomson KPB, Edwards G. On the Hausdorff distance used for the evaluation of segmentation results. Can J Remote Sensing. 1998;24(1):3–8. https://doi.org/10.1080/07038992.1998.10874685

87. Umesh Adiga PS, Chaudhuri BB. An efficient method based on watershed and rule-based merging for segmentation of 3-D histo-pathological images. Pattern Recogn. 2001;34(7):1449–58. https://doi.org/10.1016/s0031-3203(00)00076-5

88. Wright T, Klein M, Wieczorek J. A primer on visualizations for comparing populations, including the issue of overlapping confidence intervals. Am Stat. 2019;73(2):165–78. https://doi.org/10.1080/00031305.2017.1392359

89. El Jurdi R, Petitjean C, Honeine P, Cheplygina V, Abdallah F. High-level prior-based loss functions for medical image segmentation: a survey. Comput Vis Image Underst. 2021;210:103248. https://doi.org/10.1016/j.cviu.2021.103248

90. Paul A, Tang YX, Shen TC, Summers RM. Discriminative ensemble learning for few-shot chest X-ray diagnosis. Med Image Anal. 2021;68:101911. https://doi.org/10.1016/j.media.2020.101911

91. Mittal A, Hooda R, Sofat S. Lung field segmentation in chest radiographs: a historical review, current status, and expectations from deep learning. IET Image Processing. 2017;11(11):937–52. https://doi.org/10.1049/iet-ipr.2016.0526

92. Wang B, Takeda T, Sugimoto K, Zhang J, Wada S, Konishi S, et al. Automatic creation of annotations for chest radiographs based on the positional information extracted from radiographic image reports. Comput Methods Programs Biomed. 2021;209:106331. https://doi.org/10.1016/j.cmpb.2021.106331

93. Noble JA, Boukerroui D. Ultrasound image segmentation: a survey. IEEE Trans Med Imaging. 2006;25(8):987–1010. https://doi.org/10.1109/tmi.2006.877092

94. Chen H, Qi X, Yu L, Dou Q, Qin J, Heng PA. DCAN: deep contour-aware networks for object instance segmentation from histology images. Med Image Anal. 2017;36:135–46. https://doi.org/10.1016/j.media.2016.11.004

95. Heimann T, Meinzer HP. Statistical shape models for 3D medical image segmentation: a review. Med Image Anal. 2009;13(4):543–63. https://doi.org/10.1016/j.media.2009.05.004

96. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in medical imaging: a survey. Med Image Anal. 2023;88:102802. https://doi.org/10.1016/j.media.2023.102802

97. Lind Plesner L, Müller FC, Brejnebøl MW, Laustrup LC, Rasmussen F, Nielsen OW, et al. Commercially available chest radiograph AI tools for detecting airspace disease, pneumothorax, and pleural effusion. Radiology. 2023;308(3):e231236. https://doi.org/10.1148/radiol.231236

98. Afrose S, Song W, Nemeroff CB, Lu C, Yao D. Subpopulation-specific machine learning prognosis for underrepresented patients with double prioritized bias correction. Communications Medicine. 2022;2:111. https://doi.org/10.1038/s43856-022-00165-w

99. Yuan H. Toward real-world deployment of machine learning for health care: external validation, continual monitoring, and randomized clinical trials. Health Care Science. 2024;3(5):360–4. https://doi.org/10.1002/hcs2.114

100. Hu Q, de F Souza LF, Holanda GB, Alves SSA, Dos S Silva FH, Han T, et al. An effective approach for CT lung segmentation using mask region-based convolutional neural networks. Artif Intell Med. 2020;103:101792. https://doi.org/10.1016/j.artmed.2020.101792

101. Azizi S, Culp L, Freyberg J, Mustafa B, Baur S, Kornblith S, et al. Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. Nat Biomed Eng. 2023;7(6):756–79. https://doi.org/10.1038/s41551-023-01049-7

102. Mao C, Yao L, Luo Y. ImageGCN: multi-relational image graph convolutional networks for disease identification with chest X-rays. IEEE Trans Med Imaging. 2022;41(8):1990–2003. https://doi.org/10.1109/TMI.2022.3153322

How to cite this article: Yuan H, Hong C, Tran NTA, Xu X, Liu N. Leveraging anatomical constraints with uncertainty for pneumothorax segmentation. Health Care Sci. 2024;3:456–74. https://doi.org/10.1002/hcs2.119