

Received 30 July 2024, accepted 26 August 2024, date of publication 3 September 2024, date of current version 11 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3454188

## RESEARCH ARTICLE

# Automatic Lung Segmentation in Chest X-Ray Images Using SAM With Prompts From YOLO

**EBRAHIM KHALILI<sup>1,2</sup>, BLANCA PRIEGO-TORRES<sup>1,2</sup>, ANTONIO LEÓN-JIMÉNEZ<sup>1,2,3</sup>, AND DANIEL SANCHEZ-MORILLO<sup>1,2</sup>**

<sup>1</sup>Department of Engineering on Automation, Electronics and Computer Architecture and Networks, University of Cádiz, 11519 Puerto Real, Spain

<sup>2</sup>Biomedical Research and Innovation Institute of Cadiz (INiBICA), 11009 Cádiz, Spain

<sup>3</sup>Pulmonology Department, Puerta del Mar University Hospital, 11009 Cádiz, Spain

Corresponding author: Daniel Sanchez-Morillo (daniel.morillo@uca.es)

This work was supported by Consejería de Universidad, Investigación e Innovación de la Junta de Andalucía, under Grant ProyExcel\_00942.

**ABSTRACT** Despite the impressive performance of current deep learning models in the field of medical imaging, transferring the lung segmentation task in X-ray images to clinical practice is still a pending task. In this study, the performance of a fully automatic framework for lung field segmentation in chest X-ray images was evaluated. The framework is rooted in the combination of the Segment Anything Model (SAM) with prompt capabilities, and the You Only Look Once (YOLO) model to provide effective prompts. Transfer learning, loss functions, and several validation strategies were thoroughly assessed. This provided a complete benchmark that enabled future research studies to fairly compare new segmentation strategies. The results achieved demonstrated significant robustness and generalization capability against the variability in sensors, populations, disease manifestations, device processing, and imaging conditions. The proposed framework was computationally efficient, could address bias in training over multiple datasets, and had the potential to be applied across other domains and modalities.


**INDEX TERMS** Biomedical X-ray imaging, image segmentation, lung, deep learning.

## I. INTRODUCTION

Medical imaging plays a pivotal role in clinical practice and biomedical research because of its potential to provide advanced visualization of the inside of the human body, offering valuable information that contributes to early diagnosis and personalized medical care. Medical images, acquired through a variety of technologies, such as radiography, magnetic resonance imaging, computed tomography, and ultrasound, are essential tools for diagnosis, disease monitoring, computer-assisted surgery, and treatment planning [1]. The subfield of medical image segmentation is part of routine examinations. Accurate image segmentation techniques cover the automated identification and annotation of medical regions of interest (ROI) delivering critical information about the volumes or shapes of anatomic structures or medical abnormalities. Automated image segmentation is

an important step in medical image analysis and an essential stage of computer-aided diagnosis (CAD) systems [2].

In recent decades, numerous researchers have developed a wide array of automatic segmentation techniques for medical imaging, customized to the specific imaging modality and body part being studied. The first approaches were based on conventional digital image processing techniques (such as thresholding, edge detection, or region-based methods) [3]. Later, artificial intelligence, through automatic learning techniques with hand-crafted features, became a dominant approach for years [4]. These techniques based on feature engineering require a significant effort and need experts to identify the proper features to be fed into a learnable decision algorithm [5]. Currently, the rise in computational power, the availability of large datasets, and advanced model training algorithms have boosted Deep Learning (DL) as a subfield of machine learning able to replace handcrafted engineering with efficient unsupervised or semi-supervised feature learning algorithms [6]. DL takes advantage of multiple processing

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello .

layers to learn data representations with different levels of abstraction, identifying and learning suitable input features in a fully data-driven manner. This approach delivers impressive performance compared to traditional ML methods [7].

DL models can handle raw data, generalize to unseen images, and eliminate the need for handcrafted features by domain experts [8]. Besides segmentation, DL-based models have proven significant success in various medical imaging tasks, supporting clinical decision-making [9].

Lung segmentation, defined as the computer-based process of delineating lung boundaries from surrounding thoracic tissue, is a prerequisite for the automated analysis of radiological lung images to assess pulmonary lesions [10]. Despite the wide range of DL-based lung segmentation approaches, and the sophisticated pipelines developed in recent years demonstrating effectiveness in medical imaging and other domains, their clinical applicability across different pathologies remains limited [11].

Several factors contribute to the limited clinical applicability of existing lung segmentation approaches. Most current methods perform adequately when the lungs show minimal or no pathological conditions. However, in the presence of moderate to severe abnormalities such as pleural effusions, consolidations, opacities, or masses, these models often produce inaccurate segmentation, impeding the use of CAD systems in clinical settings [10].

Furthermore, developing accurate DL-based automated systems necessitates sufficient annotated data, which requires highly qualified experts and is tedious and time-consuming work [12]. The tasks of reading, labeling, and creating annotations for image masks in chest X-rays (CXRs) and computed tomography (CT) scans are exponents of this situation. Additionally, it is usual that when annotated datasets exist, they are limited to healthy subjects, individuals without severe pathology, or patients with a dominant pathology [11].

This lack of data variability leads to trained models that struggle to generalize the segmentation task beyond the specific cohorts they were trained on. This limitation hinders the effectiveness of segmentation tasks on new and unseen data, contributing to dataset bias. Dataset bias has become a major challenge in medical image segmentation [13].

In addition to dataset variability and annotation issues, there are other significant barriers. These include the diversity of medical equipment used for image acquisition and the complexities associated with sharing data across different medical centers. Legal, ethical, and privacy concerns often obstruct the seamless sharing of medical image data [14].

Despite advancements in automatic lung segmentation systems, their transfer to clinical practice remains limited. Routine image studies still heavily rely on semi-automatic segmentation methods or human inspection of automatically generated organ masks [15]. These challenges underscore the ongoing complexity of integrating advanced segmentation technologies into everyday clinical workflows.

Chest X-ray (CXR) is indeed one of the most extended medical imaging modalities to examine for pulmonary and

heart disorders the chest's anatomical structures, including the heart, lungs, blood vessels, airways, bones, and spine.

Therefore, the development of CAD systems is crucial for supporting healthcare professionals, particularly pulmonologists and radiologists, in the assessment of pulmonary and cardiac conditions.

As a core part of these systems, the task of segmenting the lung fields is critical to enable the automatic delivery of precise information about the anatomical structures identifiable in CXR images (e.g., quantification of lung nodules [16], detection of lung disease [17], or assessment of heart failure [18]).

Notwithstanding the crucial importance of lung segmentation in CXR, achieving accurate segmentation remains a significant challenge. CXR interpretation is widely acknowledged as one of the most complex tasks in radiography [19], and the development of a generic automated solution that can reliably operate in routine clinical settings without expert intervention is still pending [20].

In this context, DL-based frameworks have emerged as a promising alternative for achieving precise and clinically relevant segmentation of lung fields. Pre-trained foundational models are revolutionizing the segmentation landscape by enhancing flexibility, adaptability, and accuracy across various scenarios, often outperforming specialized models [21]. These foundational models can be fine-tuned using techniques such as prompt engineering, where prompts like points, bounding boxes, or masks guide the model to facilitate downstream tasks.

Implementing a prompt-guided foundational model framework enables the incorporation of designed data cues, enhancing the model's ability to generalize effectively to new data distributions even without prior exposure—a concept known as zero-shot generalization [22].

As a reference, the Segment Anything Model (SAM) stands out as a foundational model that has shown promising zero-shot segmentation results across various natural image datasets [23]. However, while foundational models have achieved remarkable success with regular images, their application to medical images faces significant challenges. The computational requirements pose significant barriers to the accessibility and scalability of real-world clinical applications. Visual prompt-based models, such as SAM, which utilize high-performing image encoders, demand substantial throughput processing capabilities to ensure practicality. Furthermore, foundational models in vision face general challenges related to training complexity, network architecture, privacy considerations, and bias [21]. Addressing these obstacles is crucial for deploying effective and trustworthy systems in clinical settings.

To tackle these challenges, we propose implementing a fully automatic framework for segmenting lung fields in CXR images. This framework combines the strengths of the SAM model, which excels in zero-shot segmentation with its prompt capabilities, and the You Only Look Once (YOLO) model, known for its efficient single-stage object detection

approach [24], crucial for real-time applications in clinical settings where rapid analysis is essential.

We hypothesized that the approach based on identifying bounding boxes as prompts to eliminate background noise and mitigate side effects by isolating specific areas of interest has the potential to significantly influence the model's ability to generalize across datasets of varying scales. Our research specifically delves into transfer learning methods and optimizing the process of selecting the most efficient prompts to unravel relationships between prompts and outcomes.

The main contributions of this study are as follows:

- (a) Integration of foundational models and DL algorithms. This study implements a fully automatic framework that combines SAM with the YOLO model. This combination leverages the high-performing image encoders of SAM with the prompt capabilities of YOLO, improving segmentation accuracy and generalization across different datasets.
- (b) Utilization of diverse benchmark datasets. We evaluate the framework using five benchmark datasets which include chest X-rays with varying pathologies such as COVID-19, pneumonia, and tuberculosis, enhancing the model's generalization capability and robustness.
- (c) Providing a fair benchmarking. The study reports high-performance metrics, demonstrating the effectiveness, robustness, and generalization of the proposed method given the variability in devices, imaging conditions, populations, and disease manifestations. As a result of this thorough assessment, a valuable benchmark for fair comparison in future research is provided.
- (d) Addressing clinical applicability. The study tackles the challenge of clinical applicability by focusing on transfer learning methods and optimizing the process of selecting effective prompts. This approach aims to reduce background noise and isolate specific areas of interest, showing potential to improve the model's performance in real clinical settings.

These contributions collectively advance the state of the art in lung segmentation on chest X-rays by improving accuracy, generalization, and clinical applicability through innovative model integration, preprocessing techniques, and evaluation methodologies.

The rest of the manuscript is organized as follows. In Section II, a review of existing literature on lung segmentation in CXR is provided.

Section III outlines our methodology, encompassing details on the datasets used for training and validating the DL models. It describes the preprocessing techniques applied to CXRs, the SAM architecture, the postprocessing stage, the validation strategy, the performance metrics calculated for evaluating the models' performance, and the experimental environment regarding the hardware and software resources.

After that, the experimental findings are presented and discussed in Section IV, including a comparison with

other state-of-the-art approaches and describing the study limitations.

Lastly, Section V presents the conclusions and considerations for future works.

## II. RELATED WORKS

To address the challenges and obstacles associated with lung segmentation in CXR images, a diverse range of DL-based techniques has been studied in recent years. Notably, in 2023, the MWG-UNet framework was presented [25]. MWG-UNet is based on the Wasserstein generative adversarial network U-shape network and was designed to segment the lung fields and heart in CXRs. This approach uses an attention mechanism to improve the performance of the segmentation process.

Also in 2023, the lung segmentation in CXRs was addressed using semantic segmentation and five different methods based on vision transformers. These methods used attention mechanisms and differentially weighed the significance of each part of the data input sequence [26].

In 2022, a two-stage model was proposed for lung segmentation in CXRs. The preprocessing stage included a deep learning model combining a Deep Belief Network and K-Nearest Neighbor, while the refinement stage utilized an improved principal curve method and a machine learning method [27]. In the same year, the typical U-Net architecture was adapted by replacing the convolution block with a dilated convolution block to extract multi-scale context features, each with different receptive field sizes [28].

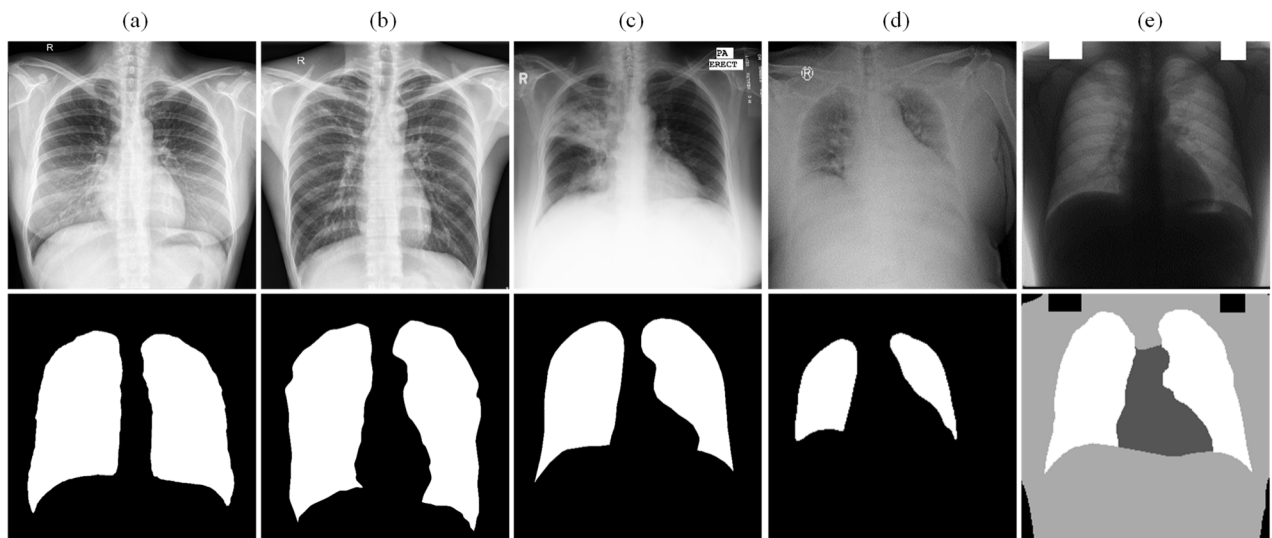
Furthermore, an unsupervised tile-wise autoencoder (T-AE) pretraining architecture was assessed for acquiring transferable knowledge, followed by fine-tuning with a U-Net segmentation model [29]. In the same year, HybridGNet, another UNet-based architecture, was introduced [30].

In 2021, a multi-scale adversarial domain adaptation network (MS-AdaNet) was introduced to enhance the cross-domain lung segmentation task, providing foundational knowledge for the field [31]. In 2019, an approach featuring two deep convolutional neural network models (an AlexNet-based CNN and a ResNet-based CNN) was evaluated to address the problem of dense abnormalities in CXRs [32].

In 2018, three fully convolutional architectures were described, including the introduction of the Inverted-Net fully convolutional network for the automated segmentation of anatomical organs in CXRs, such as lungs, clavicles, and heart [33]. Additionally, the SegNet network, with its encoder-decoder architecture, was assessed for similar tasks [34].

## III. METHODOLOGY

The approach proposed in this study included several steps. The process began with data acquisition from public datasets, followed by essential preprocessing steps such as image resizing, enhancement, and augmentation. Subsequently, the YOLO and SAM models were developed and evaluated using various validation approaches.



**FIGURE 1.** Examples of chest X-ray images extracted from the used datasets. In each case, the X-ray image is illustrated at the top, and the ground-truth mask at the bottom: (a): Darwin v7 dataset; (b): Shenzhen dataset; (c): Montgomery dataset; (d): RSUA dataset; (e): JSRT dataset.

### A. DATASETS

Five public datasets were used to train the proposed lung segmentation approach and evaluate its performance using a wide range of validation techniques.

The Montgomery dataset (MC) [35] comprises 138 posteroanterior CXRs, with 80 showing normal conditions and 58 exhibiting various degrees of tuberculosis manifestations. The images are represented in 12-bit grey-scale format and have a size of  $4020 \times 4892$  pixels.

The Shenzhen (SZ) [35] dataset consists of 662 frontal CXRs, with 326 normal cases and 336 cases showing tuberculosis manifestations. The image size varies, but the average size is around  $3000 \times 3000$  pixels.

The Japanese Society of Radiological Technology (JSRT) dataset [36] includes 247 posteroanterior CXRs with a resolution of  $2048 \times 2048$  pixels. Among these, 154 cases exhibit lung nodules, while 93 cases do not.

The RSUA dataset [37] comprises 292 CXRs with corresponding ground truth annotations by radiologists. This dataset includes 207 images from COVID-19 patients, 53 depicting pneumonia cases, and 32 showing no abnormalities. The images have a size of  $256 \times 256$  pixels.

Finally, the Darwin V7 dataset (DV7) [38] includes image heterogeneity in terms of sources, orientations, and resolutions. Image size varies from  $156 \times 156$  pixels to  $5600 \times 4700$  pixels. The DV7 dataset includes 6106 CXR images obtained from individuals diagnosed with different respiratory diseases.

All these datasets comprise CXRs obtained from patients with a diagnosis of COVID-19, pneumonia, or tuberculosis. By including images with varying pathologies and disease patterns, we aimed to enhance the model's generalization capability.

The data for training and validating the models were derived from expert-annotated images. Experts provided and

curated these annotations, ensuring the ground truth was highly reliable.

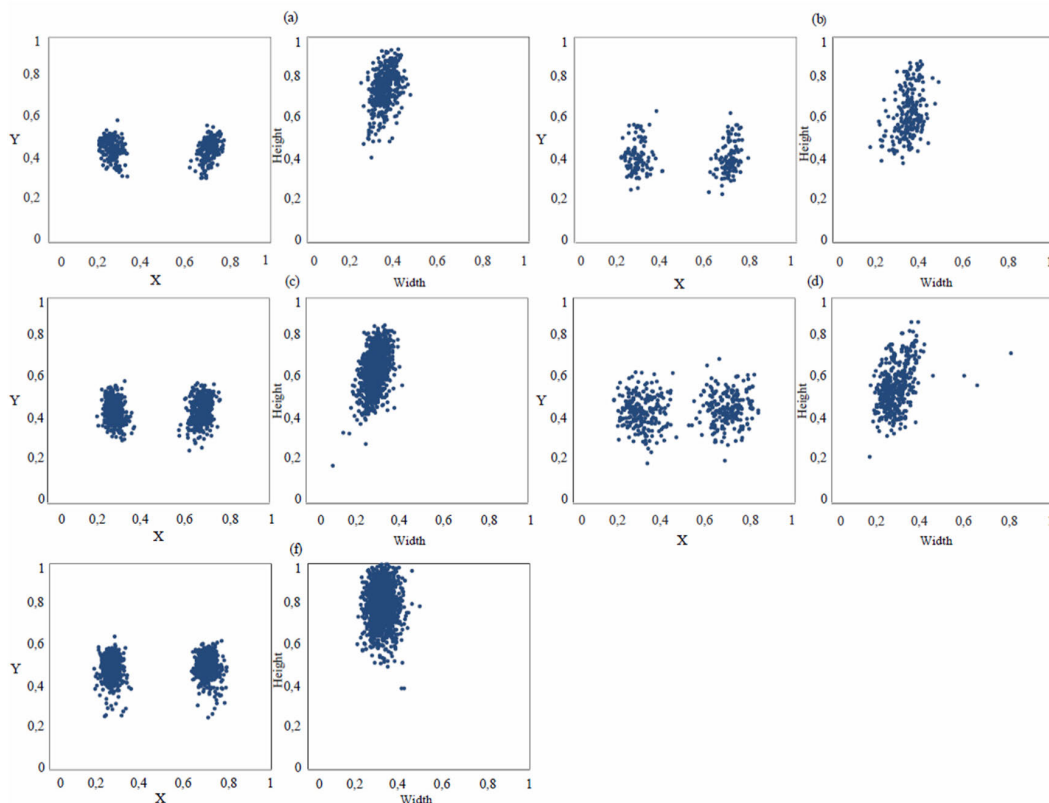
As a result, the models were built and evaluated using precise and trustworthy reference standards.

Fig. 1 represents a CXR from each dataset used in this study, along with its associated lung field mask.

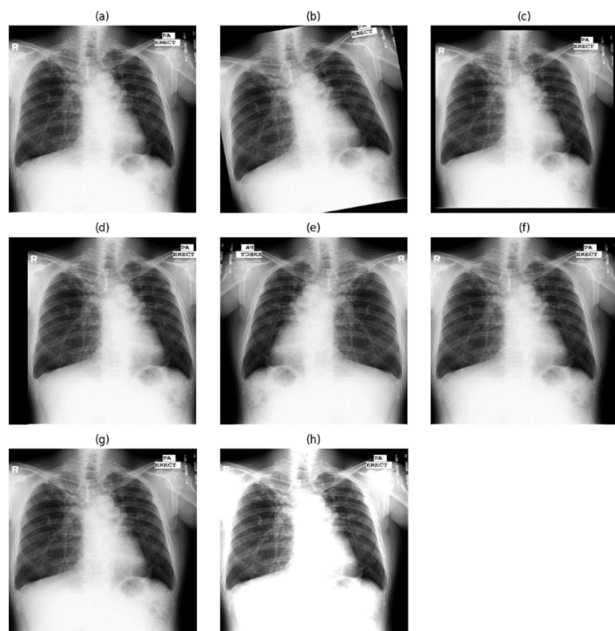
### B. DATA AUGMENTATION

Image augmentation is a key technique in DL, which involves applying transformations to original images to increase the diversity of the training datasets [39]. Due to the variation in lung sizes, as illustrated in Fig. 2, data augmentation methods were used to train and validate the YOLO model. Augmented images were dynamically generated during model training through various techniques, including rotation ( $\pm 10$  degrees), scaling ( $\pm 0.05$ ), translation (random image shifting left, right, up, or down by 10%), left-to-right flipping (with a 50% probability), and perturbation of the HSV color space (adjustment up to 1.5% of the full hue range, 70% of the saturation range, and 40% of the value range).

Scaling was used to simulate natural variation in lung sizes. Shifting images horizontally ensured that the model was exposed to different parts of the lung region during training, enhancing its ability to identify lung structures regardless of their position. Random rotations and flips were applied to simulate different orientations of the lungs, which helped the model to generalize better and improved its robustness to anatomical variability. Introducing slight variations in HSV color space simulated different imaging conditions or subtle changes in the imaging environment, which could improve the model's ability to generalize to variations encountered in real-world scenarios. Examples of augmented images are shown in Fig. 3.



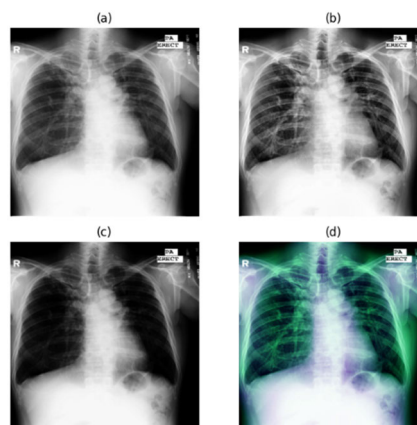
**FIGURE 2.** Distribution of the size of lung fields in the images used to develop the YOLO model. In each pair, the figure on the left illustrates the variability in size in the left and right lungs, where the coordinates of each point represent the center of the box delimiting the right and left lung fields. The figure on the right denotes, for each dataset, the width and height of the right and left bounding boxes. X and Y axes values are represented in normalized units. (a): JSRT dataset; (b): Montgomery dataset; (c): Shenzhen dataset; (d): RSUA dataset; (e): Darwin dataset.



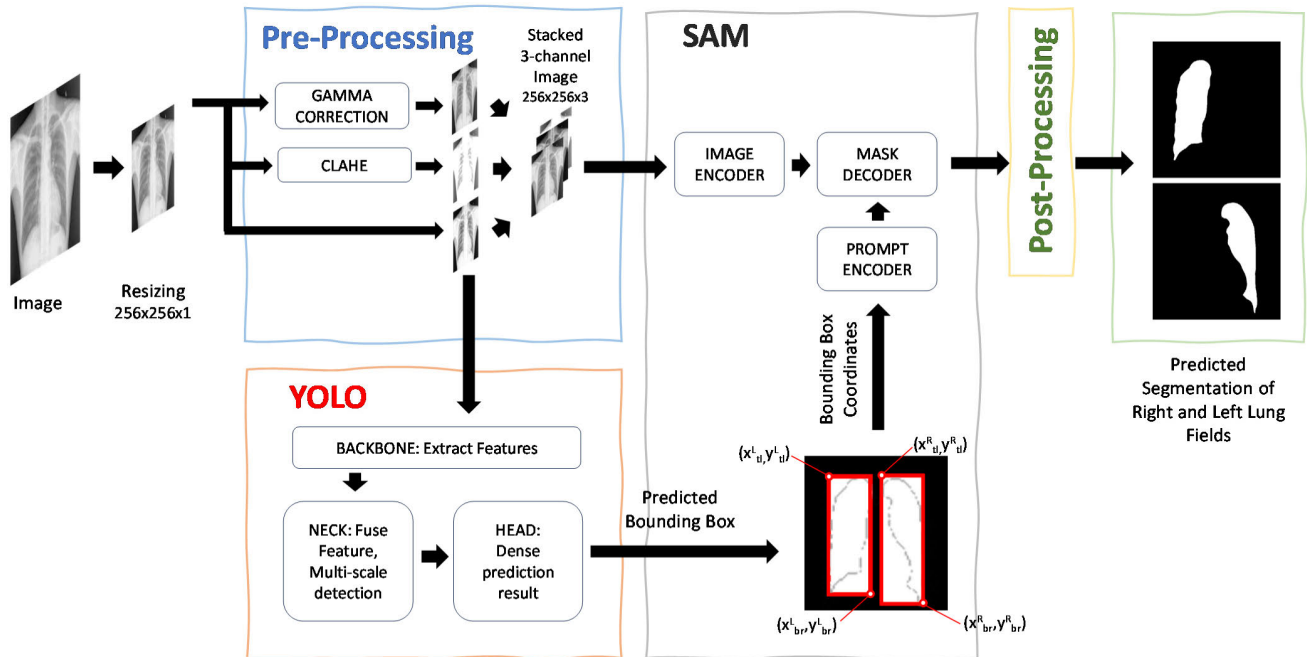
**FIGURE 3.** Examples of augmented images. (a): Original image with pixel normalization; (b): 10 degrees rotated image; (c): scaled image (x0.95); (d): shifted image (10% to the right); (e): left-to-right flipped image; (f): image with increased HSV-hue; (g) image with increased HSV-saturation; (h): image with increased HSV-value.

### C. PREPROCESSING

The images in the datasets were resized to  $256 \times 256$  pixels to standardize size, reduce memory usage, improve computational performance, and enable fair comparison with state-of-the-art approaches that commonly use this image size. Pixel intensity normalization was applied to enhance convergence during training.



**FIGURE 4.** Examples of images under different preprocessing techniques. (a): Original image with pixel normalization; (b): image preprocessed with contrast-limited adaptive histogram equalization (CLAHE); (c): image preprocessed with gamma-correction; (d): stacked 3-channel image.



**FIGURE 5.** The architecture of the proposed segmentation framework. The full-scale resolution lung radiography image is preprocessed and fed into the You Only Look Once (YOLO) network to generate the bounding boxes of lung fields. The predicted bounding boxes are used as prompts to the Segment Anything Model (SAM) model, which receives as input a stacked 3-channel image built with the gamma correction and the contrast enhancement (CLAHE) techniques. Finally, the lung boundaries provided by the SAM network are post-processed to generate the final segmentation.

Contrast-limited adaptive histogram equalization (CLAHE) was used to enhance the local contrast of images. CLAHE has proven effective in improving segmentation performance [40] by highlighting lung boundaries and enhancing the discrimination between the lungs and surrounding tissues (Fig. 4b).

Gamma correction was applied to adjust the luminance of the images. This technique allowed managing the brightness and contrast, thereby improving the visibility of the lung structures, particularly in cases with significant differences in image brightness. Gamma correction can enhance model performance in CXR classification and segmentation supporting more accurate medical image interpretation (Fig. 4c) [41].

The original normalized, CLAHE-enhanced, and gamma-corrected images were combined to build the 3-channel image (Fig. 4d), which was then input into the SAM network.

#### D. PROPOSED ARCHITECTURE

The proposed architecture, illustrated in Fig. 5, consists of an ensemble of two models: a lung fields detector based on YOLO; and a SAM-based segmentation model that delineates the lung contours.

The YOLO model performed the detection and generated prompts indicating the specific regions of interest (ROIs) in the CXR images. By accurately identifying potential lung regions, YOLO guided the SAM model to focus on these areas for more detailed segmentation.

##### 1) YOU ONLY LOOK ONCE (YOLO) MODEL

A relational module was implemented using YOLO to detect the left and right lung fields and to consider the connections between diseases and anatomical components. YOLO

processes the CXR image to detect ROIs, particularly focusing on areas likely to contain lung regions. YOLO generates bounding boxes around these ROIs, providing precise coordinates that highlight potential lung boundaries. For this study, the YOLO v5 Ultralytics version was used [42].

The normalized image is input into the YOLO backbone for feature extraction. The backbone estimates feature maps of varying sizes which are then fused in the feature fusion network (neck). These features are forwarded to the prediction head to generate multi-dimensional arrays representing bounding boxes. These bounding boxes serve as input for the SAM prompt encoder.

Bounding box coordinates are determined by fitting the smallest possible bounding box around the original segmentation masks for both lung fields. Each bounding box was defined by its top-left corner ( $x_{tl}^i, y_{tl}^i$ ) and its bottom-right corner ( $x_{br}^i, y_{br}^i$ ). The identification of the two largest contours, representing the left ( $i = L$ ) and right ( $i = R$ ) lung fields, was based on the min-max criterion detailed in Equations 1-2.

$$x_{tl}^i = \min(x | x = 1), y_{tl}^i = \min(y | y = 1) \quad (1)$$

$$x_{br}^i = \max(x | x = 1), y_{br}^i = \max(y | y = 1) \quad (2)$$

Non-maximum suppression was applied to remove redundant bounding boxes. The generalized intersection over union (GIoU) was used as a measure in the suppression algorithm. GIoU, as defined in Equation 3, was used to characterize the similarity between closely spaced detection boxes. It provides a more comprehensive description of the relative position and overlapping between two bounding boxes, A and B, compared to traditional intersection over union (IoU). To estimate GIoU, a minimum rectangle C is circumscribed. The GIoU

value is not zero if A and B do not overlap.

$$GIoU = \frac{|A \cap B|}{|A \cup B|} - \frac{|C \setminus (A \cup B)|}{|C|} \quad (3)$$

## 2) SEGMENT ANYTHING MODEL (SAM)

Bounding boxes provided by YOLO were used by SAM to focus on specific areas and perform segmentation by applying image encoding techniques and leveraging contextual information. SAM was chosen for its high-performing image encoding mechanisms, which excel at delineating lung boundaries in images even with limited annotated data, thanks to its foundation model trained on a vast dataset.

SAM has an encoder-decoder-based architecture. The decoder block uses prompt self-attention and cross-attention mechanisms, generating embeddings that focus on the object of interest [43]. Foundational models, like SAM, can extend their capabilities to encompass tasks and data distributions, allowing models to adapt beyond those encountered during their initial training. This ability, known as task generalization, allows models to adapt to new scenarios effectively [23]. Implementing this capability often involves the use of prompt engineering techniques.

In the realm of vision-language models or models driven by visual input, prompt engineering finds its primary application in two key areas: first, it facilitates the transformation of vision datasets into training data that combines both images and text, enabling human interaction with foundational models. Second, it empowers the use of vision-language models for tasks related to visual perception.

In this study, we explored how a geometric prompt can enhance the SAM model's generalization. Specifically, we considered using a bounding box as a prompt to isolate a particular ROI. This approach aimed to improve object localization by focusing on relevant areas and reducing background noise. It also sought to mitigate over-generalization biases caused by variations in capture conditions (device, image capture, cables or tubes), human operators, and anatomical variability among patients and diseases [44]. We hypothesized that this method would enhance the model's ability to generalize across different scenarios.

The SAM model was trained using a hybrid unweighted sum of the soft Jaccard Focal [45] and the Tversky [46] loss functions. These loss functions complement each other by refining the pixel-level segmentation, addressing class imbalance, and improving overlap measures.

The Tversky loss function extends the Dice loss by assigning greater importance to false negatives compared to false positives. This approach aims to balance precision and recall.

The Jaccard Focal loss function is an enhanced version of binary cross-entropy loss, specifically crafted to tackle class imbalance by prioritizing learning from challenging, misclassified examples. This loss function drives the model to focus on correctly segmenting challenging areas, which is crucial for capturing detailed lung structures across varied pathologies. The term "soft Jaccard Focal loss" refers to this

loss function's direct use of predicted probabilities, eliminating the need to threshold and convert them into a binary mask. Recent studies have expanded on this concept [47], by applying an exponent to the Dice score or using a hybrid approach that combines Dice loss with cross-entropy.

The total network loss is denoted by  $L$  and mathematically represented as illustrated in Equations 4-7:

$$L = L_{F,\gamma} + L_{T,\alpha,\beta} \quad (4)$$

$$L_{F,\gamma} = \sum_c SJ^{1/\gamma} \quad (5)$$

$$SJ = 1 - \frac{\sum_{i=1}^N p_{ic} g_{ic} + \varepsilon}{\sum_{i=1}^N p_{ic} + g_{ic} - \sum_{i=1}^N p_{ic} g_{ic} + \varepsilon} \quad (6)$$

$$L_{T,\alpha,\beta} = \frac{\sum_{i=1}^N p_{ic} g_{ic} + \varepsilon}{\sum_{i=1}^N p_{ic} g_{ic} + \alpha \sum_{i=1}^N p_{ic} \bar{g}_{ic} + \beta \sum_{i=1}^N p_{ic} g_{i\bar{c}} + \varepsilon} \quad (7)$$

where  $L_{F,\gamma}$  denotes the soft Jaccard Focal loss and  $L_{T,\alpha,\beta}$  represents the Tversky loss component, respectively.

The parameter  $\gamma$  smoothly adjusts the rate at which easy examples are downweighted, while  $\alpha$  and  $\beta$  control the trade-offs between penalizing false negatives and false positives. Here,  $g_{ic}$  represents a one-hot vector for the true labels,  $p_{ic}$  denotes a matrix containing the predicted values for each class, and the indices  $c$  and  $i$  iterate over all classes and pixels, respectively.  $N$  stands for the total number of pixels in the image. The term  $\varepsilon$  was used to prevent division by zero.

## E. POSTPROCESSING

The ROIs detected by the SAM model were post-processed to generate the final output. The Suzuki approach, which involves topological structural analysis through border following, was used to remove small, disconnected components from the prediction masks provided by the SAM model [48]. Following this, an opening morphological operation, followed by a closing operation, was applied to address false positive and false negative predictions. Both operations used a  $3 \times 3$  kernel size to refine the segmentation results.

## F. VALIDATION

Recent literature on lung segmentation reveals a wide variety of validation techniques used by researchers. However, there is a scarcity of studies that combine different validation methods to comprehensively describe the generalization capabilities of proposed models. To address this gap, our study applied four distinct validation approaches, providing a rigorous and fair test bed against which to compare the results of new lung segmentation strategies. Additionally, this methodology helps evaluate whether the proposed approach generalizes effectively across diverse data sources, data distributions, and pathologies that may influence radiologic image characteristics. These validation strategies are detailed in the next subsections.

### 1) 5-FOLD CROSS-VALIDATION

A 5-fold cross-validation (CV) scheme was used in the MC, SZ, JRST, and RSUA datasets to assess the robustness of the framework. In a 5-fold CV, the dataset is divided into five equal portions. During each iteration, four portions are used for training, while the remaining one is used for validation. The average results from these five evaluations are then calculated to estimate the model's capacity for generalization.

### 2) DATASETS INTEGRATION

In some studies, researchers have combined various datasets to create larger composite datasets, addressing potential biases inherent in the individual datasets. In this study, we applied a similar approach by merging images from the MC, SZ, JRST, RSUA, and DV-7 datasets. We then performed a random split into training and testing sets, using an 80 – 20% ratio.

### 3) CROSS-DATASET VALIDATION

In this validation strategy, 12 cross-domain combinations were evaluated: JSRT→SZ, JSRT→RSUA, JSRT→MC, SZ→MC, SZ→JRST, SZ→RSUA, MC→JRST, MC→SZ, MC→RSUA, RSUA→JSRT, RSUA→MC, and RSUA→SZ, where  $x \rightarrow y$  represents training the network on dataset  $x$  and testing it on dataset  $y$ .

### 4) SEMI-AUTOMATIC VALIDATION USING CROSS-DATASET

A semi-automated evaluation mode was defined to assess model performance using optimal bounding boxes as prompts. Instead of generating the bounding boxes using YOLO, this mode emulated the behavior of an expert who can accurately annotate the region corresponding to each lung field. Optimal bounding boxes were created using the ground-truth masks.

## G. METRICS OF PERFORMANCE

Precision, recall, accuracy, Dice score, and IoU were calculated to quantify the segmentation results. These metrics are extensively used for medical image segmentation studies.

Precision measures the proportion of correctly identified lung pixels among all pixels predicted as lung, with high precision indicating few non-lung areas misclassified as lung.

Recall gauges the model's ability to detect actual lung pixels, with high recall meaning most lung pixels are correctly identified.

Accuracy represents the overall correctness of the model's predictions, reflecting how well it distinguishes between lung and non-lung areas.

The Dice score assesses the overlap between predicted and actual lung regions, with high values indicating a close match.

IoU measures the extent of overlap between predicted and actual lung regions, with high IoU showing good alignment with the ground truth.

The pixels in the mask output by the proposed architecture can be classified into four distinct categories based

on their correspondence with the ground truth mask: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Equations 8-12 detail how the selected performance metrics can be quantitatively estimated from these categories:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \quad (11)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (12)$$

## H. EXPERIMENTAL SETTINGS

All models were implemented using the PyTorch platform. The training was conducted on an Ubuntu 20.04 system using an NVidia DGX Station, featuring 4 Tesla V-100 GPUs, and using CUDA 9.0/cuDNN 7.0 (NVIDIA Corporation).

The YOLO v5 model was trained for 200 epochs with an image size of  $256 \times 256$  pixels and a batch size of 32. The confidence and IoU thresholds were set to 0.70 and 0.60, respectively. These settings are suitable for scenarios where additional bounding boxes may overlap or enclose the primary bounding box.

Default hyperparameters were applied for all other settings. The SAM model was initialized using the pre-trained ViT-Base model. The prompt encoder was retained to handle the encoding of the bounding box prompt, and its parameters were updated during training. For the training stage, bounding box prompts were generated based on the ground-truth masks.

The learning rate was set at 0.0001 and decreased according to a scheduled learning strategy. The Adam optimizer was used for the pre-training process.

## IV. RESULTS

This section presents the results of the proposed framework for lung segmentation. First, the results of the ablation study, conducted to determine the best settings for the proposed approach, are presented. Then, the outcomes of the validation strategies used to evaluate the framework performance are presented.

### A. ABLATION STUDY

An ablation study is an analysis aimed at understanding the significance or contribution of individual factors or components within a system by methodically removing them and examining the resulting changes. In this study, the contribution of factors such as preprocessing, loss functions, and the number of bounding boxes was examined. These studies were conducted on the MC dataset to examine how



these variations impacted segmentation performance. Five-fold cross-validation was used in all cases.

### 1) CLAHE AND GAMMA CORRECTION

The impact of preprocessing techniques on lung segmentation performance was assessed. Four experiments were conducted using 5-fold cross-validation on the MC dataset to estimate segmentation metrics: a) images without preprocessing, b) CLAHE-adjusted images, c) gamma-corrected images, and d) a combination of original, CLAHE-enhanced, and gamma-corrected images. The results, detailed in Table 1, indicate that the model using both CLAHE-enhanced and gamma-corrected images, along with the original images, performed on par with the model using original and gamma-corrected images. The Dice and IoU metrics estimated for both combinations outperformed the other preprocessing techniques.

**TABLE 1. Results of the ablation study about the use of CLAHE and gamma correction using 5-fold cross-validation on the Montgomery dataset.**

CLAHE	Gamma-Correction	Precision	Recall	Accuracy	Dice	IoU
✗	✗	97.23	94.99	99.20	96.06	92.47
✗	✓	97.99	96.41	99.43	97.18	94.54
✓	✗	97.96	96.20	99.42	97.05	94.30
✓	✓	97.25	97.02	99.42	97.16	94.53

Precision, recall, accuracy, Dice c, and Intersection over Union (IoU) metrics are expressed as a percentage (%).

### 2) NUMBER OF BOUNDING BOXES

To determine the optimal number of bounding boxes, the SAM model without bounding box prompting was considered as the baseline. The improvement achieved using the YOLO model trained to provide a single ROI encompassing both lung fields was then evaluated (model M1). As a third approach, the inclusion of two bounding boxes, one for each lung field, provided by YOLO as prompts to SAM was assessed. This model (M2) was trained using the standard Dice loss function.

Additionally, YOLO was trained to provide two bounding boxes as prompts to SAM, which was trained using the described hybrid loss function (model M3).

Finally, a semi-automatic mode was explored (see section III). Two optimal bounding boxes were calculated from the ground-truth masks and fed to SAM, which was trained using the hybrid unweighted sum of the soft Jaccard Focal and the Tversky loss functions (Model M4).

Table 2 summarizes the results. It can be observed that the accuracy, Dice score, and IoU metric improved substantially from the baseline model, first with the addition of a single bounding box as a prompt (Model M1), and later with two bounding boxes delimiting each lung field (models M2 and M3). As shown, discrepancies in anatomy and lung-related diseases can result in inaccuracies in the model outcomes.

**TABLE 2. Results of the ablation study to determine the optimal number of bound boxes to be prompted to the SAM model. A 5-fold cross-validation on the Montgomery dataset was used.**

Model	Precision	Recall	Accuracy	Dice	IoU
Baseline	95.97	92.07	94.16	93.93	88.63
M1	93.01	97.69	99.04	95.22	90.97
M2	96.77	96.09	99.20	96.42	93.10
M3	97.25	97.02	99.42	97.16	94.53
M4	97.31	97.12	99.62	97.24	94.61

Precision, recall, accuracy, Dice score, and Intersection over Union (IoU) metrics are expressed as a percentage (%).

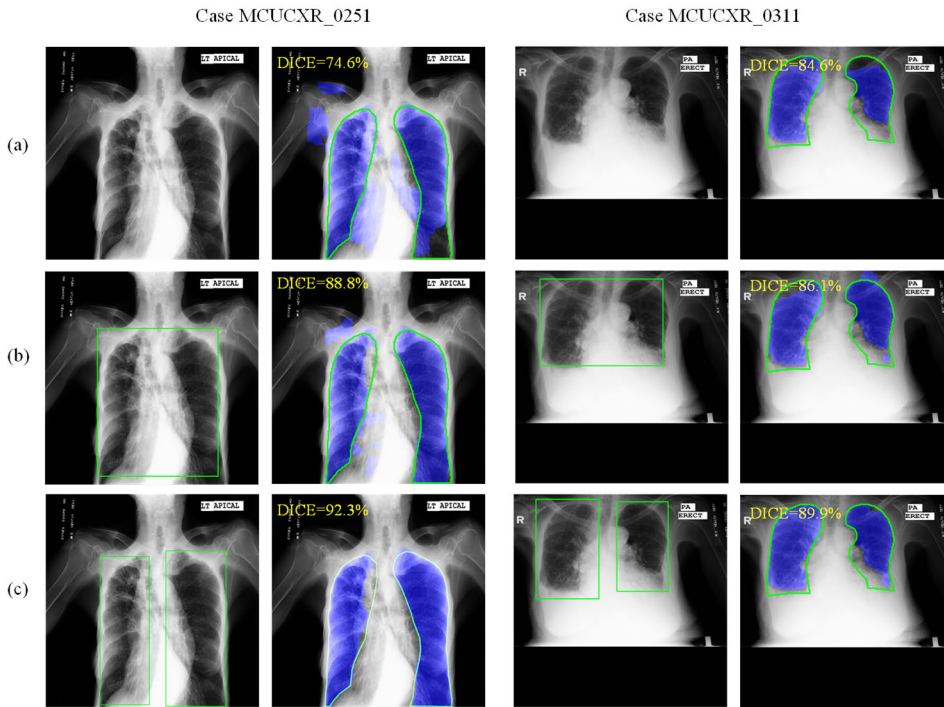
Fig. 6 illustrates how the utilization of bounding boxes can address the impact of these factors in our model. The case MCUCXR\_0251 corresponds to a 77-year-old female subject with right upper lobe fibro-cavitary disease with volume loss and tracheal deviation to the right, COPD, and scoliosis. The case MCUCXR\_0311 corresponds to an 89-year-old female subject with tuberculosis and some re-accommodation of pleural fluid along the lateral left chest wall. When there are variations in the height and width of the lung fields due to anatomical differences (case MCUCXR\_0251), the use of two bounding boxes improved the Dice score by 3.5% and 17.7% compared to using one or no bounding box, respectively.

The same was true for CXR images of cases with respiratory pathologies, which are challenging in automatic segmentation. In the case of pathological abnormalities (case MCUCXR\_0311), the Dice score increased by 3.8% using two bounding boxes compared to using only one, and by 5.3% compared to the baseline model. Therefore, it can be concluded that the combination of the YOLO and SAM models demonstrates the potential to improve the results obtained by the SAM single model and to address the main challenges in lung segmentation, such as the heterogeneity in the anatomical characteristics of these organs, and the significant differences that appear in images of healthy subjects or subjects with different lung pathologies.

### 3) LOSS FUNCTION TUNNING

Loss functions determine how neural network models compute the overall error of the residuals for each training batch. In turn, they affect how models adjust their internal weights during back-propagation. The choice of the loss function therefore directly influences the model performance. The SAM model was trained using the sum of the soft Jaccard Focal and Tversky loss functions. For the soft Jaccard Focal loss,  $\gamma = 2$  was used. This value has been found to be optimal [45].

Regarding the Tversky component, monitoring the model's performance for different combinations of  $\alpha$  and  $\beta$  in the Tversky loss function helped in selecting the best parameters to ensure that the model effectively handles class imbalance, improving recall on underrepresented classes. Typical ranges are  $\alpha \in [0.3, 0.7]$  and  $\beta \in [0.3, 0.7]$ . Consequently, the



**FIGURE 6.** Effect on the segmentation performance of using: none (a), one (b), or two (c) bounding boxes provided as prompts for the Segment Anything Model (SAM). In each case, the image on the left is the original image, including the output of the YOLO model on the border boxes of the identified lung regions; the image on the right shows the segmentation result and the Dice score, with the ground truth delineated in green.

**TABLE 3.** Results of the ablation study about the loss function tuning in the SAM model, with 5-fold cross-validation on the MC dataset.

Loss Function	Precision	Recall	Accuracy	Dice	IoU
Hausdorff	96.71	96.77	99.34	96.71	93.70
JF	96.81	96.67	99.33	96.71	93.70
Jaccard	96.11	97.13	99.31	96.57	93.46
Tversky <sup>1</sup>	97.59	96.06	99.35	96.79	93.84
Tversky <sup>2</sup>	96.76	96.85	99.34	96.78	93.80
Tversky <sup>3</sup>	96.19	97.19	99.32	96.67	93.60
<b>Tversky<sup>1</sup> + JF</b>	<b>97.25</b>	<b>97.02</b>	<b>99.42</b>	<b>97.16</b>	<b>94.53</b>
Tversky <sup>2</sup> + JF	96.62	96.82	99.33	96.70	93.65
Tversky <sup>3</sup> + JF	96.63	96.85	99.34	96.72	93.70

Precision, recall, accuracy, Dice score, and Intersection over Union (IoU) metrics are expressed as a percentage (%). JF: soft Jaccard focal with  $\gamma = 2$ . Tversky<sup>1</sup>: Tversky function with  $\alpha = 0.3$  and  $\beta = 0.7$ . Tversky<sup>2</sup>: Tversky function with  $\alpha = 0.5$  and  $\beta = 0.5$ . Tversky<sup>3</sup>: Tversky function with  $\alpha = 0.7$  and  $\beta = 0.3$ .

following steps were taken for Tversky parameter tuning. First, the framework was trained using commonly used loss functions such as Hausdorff distance, Jaccard Focal ( $\gamma = 2$ ), and Jaccard. Then, models with different  $\alpha$  and  $\beta$  combinations were evaluated for both Tversky alone and the hybrid loss function. A 5-fold CV on the MC dataset was used in all cases. The performance of each of these combinations was logged to track metrics over epochs to ensure stable convergence. The results are shown in Table 3. The loss function that yielded the highest validation performance used the sum of the soft Jaccard Focal function with  $\gamma = 2$  and the Tversky function with  $\alpha = 0.7$  and  $\beta = 0.3$ .

#### 4) FINAL CONFIGURATION AFTER THE ABLATION STUDY

The settings estimated to provide the best results according to the ablation study were established. The image size was set to  $256 \times 256$  pixels. The ablation study confirmed the contribution of CLAHE and gamma correction to improve the segmentation results. In addition, the benefits of using two bounding boxes were validated. Finally, the hybrid unweighted sum of the soft Jaccard Focal and the Tversky functions was selected as the loss function given that it provided the best result in the performed tests.

#### B. PERFORMANCE EVALUATION

The performance of the proposed SAM-YOLO framework was evaluated across the various datasets and validation strategies described in Section III. The results of the 5-fold cross-validation are shown in Table 4 for each dataset.

**TABLE 4.** Performance of the method based on SAM with prompts from YOLO estimated using 5-fold cross-validation.

Dataset	Precision	Recall	Accuracy	Dice	IoU
JRST $\rightarrow$ JRST	97.24	97.23	99.25	97.19	94.57
MC $\rightarrow$ MC	97.25	97.02	99.42	97.16	94.53
SZ $\rightarrow$ SZ	95.56	95.16	98.97	95.25	91.07
RSUA $\rightarrow$ RSUA	97.09	96.51	99.47	96.72	93.72

Precision, recall, accuracy, Dice score, and Intersection over Union (IoU) metrics are expressed as a percentage (%). "x  $\rightarrow$  y" represents training the network on the x dataset and subsequently testing it directly on the y dataset.

Table 5 presents a comparison of the generalization performance achieved by our model across different datasets (dataset integration).

**TABLE 5. Performance of the method based on SAM with prompts from YOLO estimated using validation with dataset integration.**

Dataset	Precision	Recall	Accuracy	Dice	IoU
SZ+MC	95.80	96.91	99.22	96.27	92.94
JSRT+SZ+MC	96.98	96.80	99.25	96.70	93.67
JSRT+SZ+MC+RSUA+ +DARWIN	95.24	95.87	98.76	95.46	91.42

Precision, recall, accuracy, Dice score, and Intersection over Union (IoU) metrics are expressed as a percentage (%). "x → y" represents training the network on the x dataset and subsequently testing it directly on the y dataset.

Table 6 summarizes the model performance estimated using cross-dataset validation.

**TABLE 6. Performance of the method based on SAM with prompts from YOLO estimated using cross-datasets validation.**

Dataset	Precision	Recall	Accuracy	Dice	IoU
JRST → SZ	91.32	95.84	98.58	93.25	88.10
JRST → RSUA	96.33	93.69	98.25	89.26	81.98
JRST → MC	93.01	97.68	99.04	95.22	90.97
SZ → MC	95.24	96.76	99.18	95.90	92.21
SZ → JSRT	96.59	95.84	98.96	96.16	92.67
SZ → RSUA	90.09	91.49	98.41	90.11	83.25
MC → JSRT	97.61	94.50	98.93	96.00	92.34
MC → SZ	93.80	94.36	98.72	93.85	89.15
MC → RSUA	84.39	90.20	97.74	86.45	78.77
RSUA → JSRT	97.62	95.61	99.10	96.58	93.42
RSUA → MC	94.46	97.31	99.13	95.82	92.06
RSUA → SZ	94.51	96.22	98.96	95.24	91.06

Precision, recall, accuracy, Dice score, and Intersection over Union (IoU) metrics are expressed as a percentage (%). "x → y" represents training the network on the x dataset and subsequently testing it directly on the y dataset

Results for the semi-automatic validation using a 5-fold CV in the MC dataset are shown in Table 2 (model M4). Table 7 shows the results calculated using cross-dataset validation. The results highlight the sensitivity of the framework to errors in the lung field detection by YOLO. Both in the cross-validation on the MC dataset and, more importantly, in the cross-validation on cross-datasets, the overall segmentation performance improved when the accuracy of the contour delineation used as a prompt for SAM became higher. More specifically, in unbalanced sets, such as the case of the cross-dataset validation in MC → RSUA, the improvement in the Dice score was 7.51%.

Fig. 7 illustrates some segmentation results according to the parameters defined after the ablation study, including good and poor cases in each of the datasets and the estimated Dice score. The integration of YOLO and SAM, and the use of two bounding boxes as prompts to SAM, can produce excellent segmentation results. In some cases, the segmentation deteriorates due to unavoidable prediction errors. Some of the factors causing this deterioration include a vague shape of the lung region due to consolidation, changes in lung texture caused by the disease, and scattered white pixels [49].

**TABLE 7. Performance of the method based on SAM with prompts from YOLO semiautomatic segmentation and cross-datasets.**

Dataset	Precision	Recall	Accuracy	Dice	IoU
JRST → SZ	94.29	96.55	98.98	95.32	91.17
JRST → RSUA	92.64	94.37	98.99	93.36	87.84
JRST → MC	92.32	98.33	99.02	95.20	90.89
SZ → MC	94.66	97.78	99.21	96.14	92.63
SZ → JSRT	97.44	95.78	99.09	96.57	93.42
SZ → RSUA	96.75	90.99	99.05	93.62	88.24
MC → JSRT	98.39	94.54	99.05	96.41	93.10
MC → SZ	96.39	95.04	99.04	95.61	91.70
MC → RSUA	95.80	82.36	99.07	93.96	88.74
RSUA → JSRT	97.82	96.41	99.11	97.10	94.38
RSUA → MC	94.37	98.14	99.21	96.20	92.70
RSUA → SZ	95.37	96.46	99.08	95.83	92.09

Precision, recall, accuracy, Dice score, and Intersection over Union (IoU) metrics, are expressed as a percentage (%). "x → y" represents training the network on the x dataset and subsequently testing it directly on the y dataset.

As illustrated in Fig. 7, in some cases, the chosen prompt led to segmentation failures, which suggests that a semi-automatic strategy involving an expert collaboration may be an excellent complement to the automatic mode in particularly challenging cases (e.g., case CXR\_Image\_150 shown in Fig.7).

## V. DISCUSSION

### A. COMPARISON WITH EXISTING METHODS

In this study, a fully automatic framework for lung segmentation in CXR images was evaluated.

Table 8 summarizes the methods and results of some recent works on lung segmentation, sorted by work and year of publication, highlighting the datasets used, the validation strategies, algorithms, and performance metrics. The lack of standardization in validation strategies and evaluation metrics complicates performance comparison among different methods.

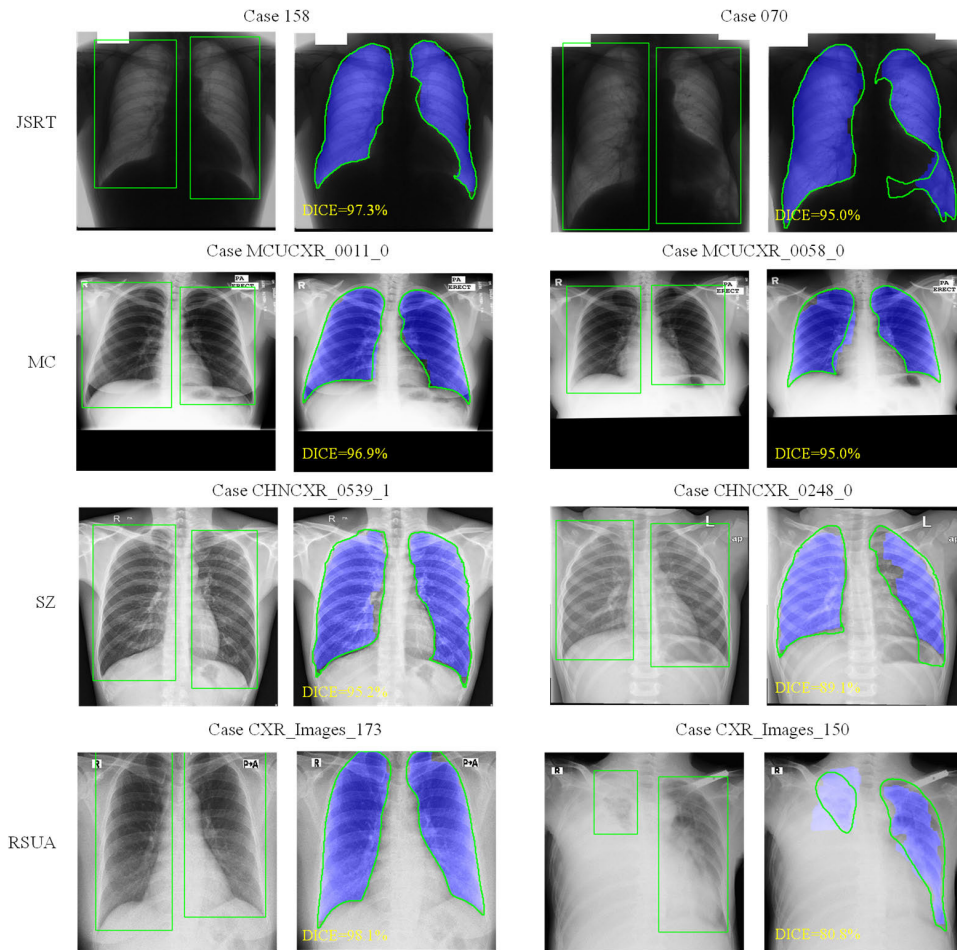
For approaches using the same dataset for training and validation, our model outperformed others on the JRST dataset with a 97.2% Dice score, compared to a DBN-CPL Hybrid (96.7% Dice score) [33] and a SegNet model (95.7% Dice score) [40].

Although the performance of the proposed approach was slightly lower than the InvertedNet CNN (97.6% Dice Score and 95.5% IoU) [39], it was still competitive.

On the MC dataset, the proposed model exceeded the performance of the AlexNet+ResNet-based CNNs presented in [38] by more than 3.5% in Dice and 6.4% in IoU metrics.

Some studies were built on a dataset augmented by integrating several individual datasets. The JRST+SZ datasets were used in [31] and [34] to build a UNet and a DCI-UNet model, respectively. Dice and IoU metrics were 95.3%, 92.3%, and 95.4%, 92.2% respectively.

The YOLO+SAM approach outperformed these results using the JRST+SZ+MC datasets and reaching a 96.7% Dice score estimated using 5-fold cross-validation. In addition, the presented approach competed well with the TranMt model, which was evaluated in [32] using hold-out validation



**FIGURE 7.** Instances of good (left) and poor segmentation (right) in the evaluated datasets: JSRT dataset; MC, Montgomery dataset; SZ, Shenzhen dataset; RSUA dataset. In each case, two images are shown: on the left side, the original image including the output of the YOLO model with the bounding boxes of the identified lung regions; on the right, the segmentation result and its Dice metric, with the ground truth delineated in green.

**TABLE 8.** Summary of the performance of lung segmentation models in chest X-rays achieved in recent studies.

Work	Year	Dataset	Validation	Model	Recall	Accuracy	Dice	IoU
[31]	2023	JRST+SZ → JRST+SZ	70-30%	MWG-UNet			95.28	92.31
[32]	2023	JRST+SZ+MC → JRST+SZ+MC	78-22%	TransMt		98.36	96.80	
[33]	2022	JRST → JRST	70-30%	DBN-CPL Hybrid			96.70	
[34]	2022	JRST+MC → JRST+MC	5-Fold CV	DCI-UNet			95.40	92.16
[35]	2022	SZ → MC	Cross-Datasets	TAE-Seg			92.45	
		MC → SZ	Cross-Datasets	TAE-Seg			95.50	
[36]	2022	JSRT → MC	Cross-Datasets	HybridGNet 2			95.40	
[37]	2021	SZ → MC	80-20%	MS-AdaNet			95.80	
		MC → SZ	Cross-Datasets	MS-AdaNet			94.10	
[38]	2019	MC → MC	78-22%	AlexNet+ResNet- based CNNs	97.54	96.97	93.56	88.07
[39]	2018	JSRT → JSRT	3-Fold CV	InvertedNet CNN			97.40	94.90
[40]	2018	JSRT → JSRT	77-23%	SegNet			95.70	

Precision, recall, accuracy, Dice score, and Intersection over Union (IoU) metrics are expressed as a percentage (%). "x → y" represents training the network on the x dataset and subsequently testing it directly on the y dataset.

(78-22%) in the same augmented dataset and reached a 96.8% Dice score.

In cross-dataset validation, the YOLO+SAM model outperformed the TAE-Seg [35] and MS-AdaNet [37] models, achieving a 96.2% Dice score compared to 92.5% and

95.8%, respectively, when training on the SZ dataset and validating on the MC dataset. Similarly, when trained on the MC dataset and validated on the SZ dataset, the proposed approach achieved a 95.6% Dice score, outperforming the other models.

Finally, the HybridGNet 2 architecture [36], assessed using the JRST dataset for training and the MC dataset for validation achieved a 95.5% Dice score. The YOLO-SAM model performance was close, with a 95.2% Dice score in the same setting.

### **B. GENERALIZATION, ROBUSTNESS, AND COMPUTATIONAL PERFORMANCE**

The comparative analysis reveals that our model delivered competitive efficiency results and exhibited a significant level of generalization ability.

The framework integrated SAM and YOLO to leverage their respective strengths. YOLO first detected and proposed ROIs in the CXR images. These ROIs were then passed to SAM, which performed detailed segmentation focusing on specific areas, enhancing precision and reducing background noise. This collaborative approach ensured that the segmentation process was both accurate and efficient, utilizing YOLO's fast detection capabilities and SAM's detailed segmentation abilities. By combining these models, the framework achieved a high level of accuracy and generalization in lung segmentation across different datasets, enhancing its potential for clinical application.

Traditionally, published studies have used only one or two validation strategies, making it difficult to fairly compare results generated from different techniques. This study facilitates this task in future experiments.

This study has faced factors that likely contribute to differences in segmentation performance. The variability in lung pathologies across datasets may cause the model to perform well on healthy lungs but struggle with diseased lungs. Class imbalance can skew the model's learning, leading to biased predictions. Integrating datasets from different sources can introduce heterogeneity in terms of image quality, resolution, and acquisition protocols. This heterogeneity can affect the model's ability to generalize across different types of data. Differences in the distribution of images from different datasets (e.g., variations in patient demographics, equipment used, and imaging settings) can lead to domain shifts, where the model performs well on one dataset but poorly on another. The amount of training data available for each validation strategy can impact performance. More data generally leads to better generalization.

It must be noted that the global model's effectiveness heavily depends on the accuracy of the bounding boxes generated by YOLO. Inaccurate lung region detection by YOLO can degrade performance. All these factors have been addressed by extending the validation to a heterogeneity of datasets and validation strategies. The results demonstrate the robustness and generalization of the lung segmentation task, given the variability in sensors, populations, disease manifestations, device processing, and imaging conditions. The results are promising and contribute to the research of segmentation models for clinical routine.

Additionally, the combination of YOLO and SAM can provide a more computationally efficient solution compared

to encoder-decoder models. By using YOLO's efficient detection followed by SAM's focused segmentation, highly competitive results were obtained with less computational burden. YOLO processes images faster than many other object detection models due to its single-stage detection architecture. When using YOLO to narrow down the areas to segment, the SAM model can focus on a smaller region, reducing the overall computational load compared to segmenting the entire image. This efficiency is particularly beneficial in clinical settings where quick turnaround times are essential, such as in emergency departments or routine radiographic screenings.

### **C. BIAS**

Differences in patient populations, imaging techniques, and pathologies across datasets can introduce biases that affect performance. Publicly available datasets used in this study include images with a high degree of heterogeneity in terms of patient country of origin, hospital center, acquisition devices, and pathologies.

Bias in CXRs can originate from various factors. Firstly, the diversity in anatomical structures can result in variations in the location of points of interest. Secondly, pathologies can lead to alterations in the texture of the lungs, which may, in turn, create variations in the contrast observed in X-ray images.

Additionally, human experts consider the entire target region and opt for a smooth transition during the segmentation process, while machine learning methods examine individual pixels and their interconnections to determine the most suitable match for the target area. This distinction leads to variations in decisions at the image edges.

Fig. 2b illustrates how the RSUA dataset has significant variation in lung sizes, potentially leading to model bias, which can be relevant for cross-dataset validation. As appreciated in Fig. 7, the proposed combination of YOLO and SAM has the potential to identify and locate the volumes and shapes of common anatomical structures in the different datasets, and consequently, address bias.

### **D. LIMITATIONS**

Despite the excellent performance achieved by the framework in all tested settings, the study has some limitations. The primary limitation is that the radiographic images used are posteroanterior (PA) and anteroposterior (AP) views of CXRs; the side view is not included. Additionally, the described method only performs segmentation. Diagnosis and disease staging of respiratory pathologies will be the focus of our next study.

## **VI. CONCLUSION**

This work introduced a fully automated deep-learning framework to increase the accuracy of lung segmentation in CXRs. The framework is based on the development of an automated model capable of training on small datasets while maintaining applicability to larger and more diverse datasets.

By using a foundational model (SAM) for automatic lung segmentation in CXRs and utilizing fully automated prompts from a YOLO model, the lung segmentation goals were effectively addressed. This approach emphasizes the importance of focusing on the lung region to achieve generalization capabilities and suggests potential applications across various domains and modalities.

The framework's performance and robustness were rigorously evaluated using a range of datasets and validation methods, establishing a valuable benchmark for fair comparative analysis in future research. The results underscore the potential of novel approaches for lung segmentation, which provide robust and reliable methods with clinical applicability.

Future research will extend this segmentation task to a comprehensive, automated diagnostic pipeline, incorporating additional stages for detecting and quantifying abnormalities and tracking the progression of specific respiratory pathologies.

## REFERENCES

- [1] F. Ritter, T. Boskamp, A. Homeyer, H. Laue, M. Schwier, F. Link, and H.-O. Peitgen, "Medical image analysis," *IEEE Pulse*, vol. 2, no. 6, pp. 60–70, Nov. 2011, doi: [10.1109/MPUL.2011.942929](https://doi.org/10.1109/MPUL.2011.942929).
- [2] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proc. IEEE*, vol. 109, no. 5, pp. 820–838, May 2021, doi: [10.1109/JPROC.2021.3054390](https://doi.org/10.1109/JPROC.2021.3054390).
- [3] K. K. D. Ramesh, G. Kumar, K. Swapna, D. Datta, and S. Rajest, "A review of medical image segmentation algorithms," *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 7, Jul. 2018, Art. no. 169184, doi: [10.4108/eai.12-4-2021.169184](https://doi.org/10.4108/eai.12-4-2021.169184).
- [4] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digit. Imag.*, vol. 32, no. 4, pp. 582–596, Aug. 2019, doi: [10.1007/s10278-019-00227-x](https://doi.org/10.1007/s10278-019-00227-x).
- [5] A. S. Panayides, A. Amini, N. D. Filipovic, A. Sharma, S. A. Tsafaris, A. Young, D. Foran, N. Do, S. Golemati, T. Kurc, K. Huang, K. S. Nikita, B. P. Veasey, M. Zervakis, J. H. Saltz, and C. S. Pattichis, "AI in medical imaging informatics: Current challenges and future directions," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 1837–1857, Jul. 2020, doi: [10.1109/JBHI.2020.2991043](https://doi.org/10.1109/JBHI.2020.2991043).
- [6] C. Shen, D. Nguyen, Z. Zhou, S. B. Jiang, B. Dong, and X. Jia, "An introduction to deep learning in medical physics: Advantages, potential, and challenges," *Phys. Med. Biol.*, vol. 65, no. 5, Mar. 2020, Art. no. 05TR01, doi: [10.1088/1361-6560/ab6f51](https://doi.org/10.1088/1361-6560/ab6f51).
- [7] J. Hofmanner, F. Prayer, J. Pan, S. Röhrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *Eur. Radiol. Experim.*, vol. 4, no. 1, p. 50, Aug. 2020, doi: [10.1186/s41747-020-00173-2](https://doi.org/10.1186/s41747-020-00173-2).
- [8] L. Cai, J. Gao, and D. Zhao, "A review of the application of deep learning in medical image classification and segmentation," *Ann. Transl. Med.*, vol. 8, no. 11, p. 713, Jun. 2020, doi: [10.21037/atm.2020.02.44](https://doi.org/10.21037/atm.2020.02.44).
- [9] I. Qureshi, J. Yan, Q. Abbas, K. Shaheed, A. B. Riaz, A. Wahid, M. W. J. Khan, and P. Szczuko, "Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends," *Inf. Fusion*, vol. 90, pp. 316–352, Feb. 2023, doi: [10.1016/j.inffus.2022.09.031](https://doi.org/10.1016/j.inffus.2022.09.031).
- [10] I. H. Sarker, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," *Social Netw. Comput. Sci.*, vol. 2, no. 6, p. 420, Aug. 2021, doi: [10.1007/s42979-021-00815-1](https://doi.org/10.1007/s42979-021-00815-1).
- [11] A. Mansoor, U. Bagci, B. Foster, Z. Xu, G. Z. Papadakis, L. R. Folio, J. K. Udupa, and D. J. Mollura, "Segmentation and image analysis of abnormal lungs at CT: Current approaches, challenges, and future trends," *RadioGraphics*, vol. 35, no. 4, pp. 1056–1076, Jul. 2015, doi: [10.1148/rg.2015140232](https://doi.org/10.1148/rg.2015140232).
- [12] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Rev. Cancer*, vol. 18, no. 8, pp. 500–510, Aug. 2018, doi: [10.1038/s41568-018-0016-5](https://doi.org/10.1038/s41568-018-0016-5).
- [13] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: Methodological failures and recommendations for the future," *NPJ Digit. Med.*, vol. 5, no. 1, p. 48, Apr. 2022, doi: [10.1038/s41746-022-00592-y](https://doi.org/10.1038/s41746-022-00592-y).
- [14] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu, and Z. Xu, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2531–2540, Jul. 2020, doi: [10.1109/TMI.2020.2973595](https://doi.org/10.1109/TMI.2020.2973595).
- [15] J. M. Stein, L. L. Walkup, A. S. Brody, R. J. Fleck, and J. C. Woods, "Quantitative CT characterization of pediatric lung development using routine clinical imaging," *Pediatric Radiol.*, vol. 46, no. 13, pp. 1804–1812, Dec. 2016, doi: [10.1007/s00247-016-3686-8](https://doi.org/10.1007/s00247-016-3686-8).
- [16] H.-Y. Chiu, R. H.-T. Peng, Y.-C. Lin, T.-W. Wang, Y.-X. Yang, Y.-Y. Chen, M.-H. Wu, T.-H. Shiao, H.-S. Chao, Y.-M. Chen, and Y.-T. Wu, "Artificial intelligence for early detection of chest nodules in X-ray images," *Biomedicines*, vol. 10, no. 11, p. 2839, Nov. 2022, doi: [10.3390/biomedicines10112839](https://doi.org/10.3390/biomedicines10112839).
- [17] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, "High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images," *Comput. Biol. Med.*, vol. 155, Mar. 2023, Art. no. 106646, doi: [10.1016/j.compbiomed.2023.106646](https://doi.org/10.1016/j.compbiomed.2023.106646).
- [18] A. Celik, A. O. Surlmeli, M. Demir, K. Esen, O. Fural, and A. Camsari, "The early diagnostic value of chest X-ray scanning by the help of artificial intelligence in heart failure (Art-in-HF): The first outcomes," *J. Amer. College Cardiol.*, vol. 79, no. 9, p. 395, Mar. 2022, doi: [10.1016/s0735-1097\(22\)01386-9](https://doi.org/10.1016/s0735-1097(22)01386-9).
- [19] L. Delrue, R. Gosselin, B. Ilsen, A. Van Landeghem, J. de Mey, and P. Duyck, "Difficulties in the interpretation of chest radiography," in *Comparative Interpretation of CT and Standard Radiography of the Chest*, E. E. Coche, B. Ghaye, J. de Mey, and P. Duyck, Eds., Berlin, Germany: Springer, 2011, pp. 27–49, doi: [10.1007/978-3-540-79942-9\\_2](https://doi.org/10.1007/978-3-540-79942-9_2).
- [20] A. Mansoor, U. Bagci, Z. Xu, B. Foster, K. N. Olivier, J. M. Elinoff, A. F. Suffredini, J. K. Udupa, and D. J. Mollura, "A generic approach to pathological lung segmentation," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2293–2310, Dec. 2014, doi: [10.1109/TMI.2014.2337057](https://doi.org/10.1109/TMI.2014.2337057).
- [21] J. Ma and B. Wang, "Towards foundation models of biological image segmentation," *Nature Methods*, vol. 20, no. 7, pp. 953–955, Jul. 2023, doi: [10.1038/s41592-023-01885-0](https://doi.org/10.1038/s41592-023-01885-0).
- [22] R. Bommasani et al., "On the opportunities and risks of foundation models," 2022, *arXiv:2108.07258*.
- [23] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Paris, France, Oct. 2023, pp. 3992–4003, doi: [10.1109/iccv51070.2023.00371](https://doi.org/10.1109/iccv51070.2023.00371).
- [24] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525, doi: [10.1109/CVPR.2017.690](https://doi.org/10.1109/CVPR.2017.690).
- [25] Y. Lyu and X. Tian, "MWG-UNet: Hybrid deep learning framework for lung fields and heart segmentation in chest X-ray images," *Bioengineering*, vol. 10, no. 9, p. 1091, Sep. 2023, doi: [10.3390/bioengineering10091091](https://doi.org/10.3390/bioengineering10091091).
- [26] R. Ghali and M. A. Akhloufi, "Vision transformers for lung segmentation on CXR images," *Social Netw. Comput. Sci.*, vol. 4, no. 4, p. 414, May 2023, doi: [10.1007/s42979-023-01848-4](https://doi.org/10.1007/s42979-023-01848-4).
- [27] T. Peng, T. C. Xu, Y. Wang, and F. Li, "Deep belief network and closed polygonal line for lung segmentation in chest radiographs," *Comput. J.*, vol. 65, no. 5, pp. 1107–1128, May 2022, doi: [10.1093/comjnl/bxaa148](https://doi.org/10.1093/comjnl/bxaa148).
- [28] Md. S. Alam, D. Wang, Q. Liao, and A. Sowmya, "A multi-scale context aware attention model for medical image segmentation," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 8, pp. 3731–3739, Aug. 2023, doi: [10.1109/JBHI.2022.3227540](https://doi.org/10.1109/JBHI.2022.3227540).
- [29] Y. Chen, H. Zhang, Y. Wang, L. Liu, Q. M. J. Wu, and Y. Yang, "TAE-seg: Generalized lung segmentation via tilewise AutoEncoder enhanced network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022, doi: [10.1109/TIM.2022.3217870](https://doi.org/10.1109/TIM.2022.3217870).

- [30] N. Gaggion, L. Mansilla, C. Mosquera, D. H. Milone, and E. Ferrante, "Improving anatomical plausibility in medical image segmentation via hybrid graph neural networks: Applications to chest X-ray analysis," *IEEE Trans. Med. Imag.*, vol. 42, no. 2, pp. 546–556, Feb. 2023, doi: [10.1109/TMI.2022.3224660](https://doi.org/10.1109/TMI.2022.3224660).
- [31] J. An, Q. Cai, Z. Qu, and Z. Gao, "COVID-19 screening in chest X-ray images using lung region priors," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 11, pp. 4119–4127, Nov. 2021, doi: [10.1109/JBHI.2021.3104629](https://doi.org/10.1109/JBHI.2021.3104629).
- [32] J. C. Souza, J. O. B. Diniz, J. L. Ferreira, G. L. F. da Silva, A. C. Silva, and A. C. de Paiva, "An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks," *Comput. Methods Programs Biomed.*, vol. 177, pp. 285–296, Aug. 2019, doi: [10.1016/j.cmpb.2019.06.005](https://doi.org/10.1016/j.cmpb.2019.06.005).
- [33] A. A. Novikov, D. Lenis, D. Major, J. Hladuvka, M. Wimmer, and K. Bühler, "Fully convolutional architectures for multiclass segmentation in chest radiographs," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1865–1876, Aug. 2018, doi: [10.1109/TMI.2018.2806086](https://doi.org/10.1109/TMI.2018.2806086).
- [34] L. Saïdy and C.-C. Lee, "Chest X-ray image segmentation using encoder-decoder convolutional network," in *Proc. IEEE Int. Conf. Consum. Electron.-Taiwan (ICCE-TW)*, May 2018, pp. 1–2, doi: [10.1109/ICCE-China.2018.8448537](https://doi.org/10.1109/ICCE-China.2018.8448537).
- [35] V. Danilov, A. Proutski, A. Kirpich, D. Litmanovich, and Y. Gankin, "Chest X-ray dataset for lung segmentation," Mendeley Data, V2, 2022, doi: [10.17632/8gf9vpkhgy.1](https://doi.org/10.17632/8gf9vpkhgy.1). [Online]. Available: <https://data.mendeley.com/datasets/8gf9vpkhgy/1>
- [36] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wang, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantum Imag. Med Surg.*, vol. 4, no. 6, pp. 475–477, Dec. 2014, doi: [10.3978/j.issn.2223-4292.2014.11.20](https://doi.org/10.3978/j.issn.2223-4292.2014.11.20).
- [37] J. Shiraiishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-I. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "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," *Amer. J. Roentgenol.*, vol. 174, no. 1, pp. 71–74, Jan. 2000, doi: [10.2214/ajr.174.1.1740071](https://doi.org/10.2214/ajr.174.1.1740071).
- [38] R. Rsua, "RSUA chest X-ray dataset," Mendeley Data, V1, 2023, doi: [10.17632/2jg8vfdmpm.1](https://doi.org/10.17632/2jg8vfdmpm.1). [Online]. Available: <https://data.mendeley.com/datasets/2jg8vfdmpm/1>
- [39] E. Goceri, "Medical image data augmentation: Techniques, comparisons and interpretations," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 12561–12605, Nov. 2023, doi: [10.1007/s10462-023-10453-z](https://doi.org/10.1007/s10462-023-10453-z).
- [40] Y. Wang, Y. Guo, Z. Wang, L. Yu, Y. Yan, and Z. Gu, "Enhancing semantic segmentation in chest X-ray images through image preprocessing: PS-KDE for pixel-wise substitution by kernel density estimation," *PLoS ONE*, vol. 19, no. 6, Jun. 2024, Art. no. e0299623, doi: [10.1371/journal.pone.0299623](https://doi.org/10.1371/journal.pone.0299623).
- [41] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan, and M. E. H. Chowdhury, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput. Biol. Med.*, vol. 132, May 2021, Art. no. 104319, doi: [10.1016/j.compbiomed.2021.104319](https://doi.org/10.1016/j.compbiomed.2021.104319).
- [42] G. Jocher. (May 2020). *YOLOv5 by Ultralytics*. [Online]. Available: <https://github.com/ultralytics/yolov5>
- [43] S. Roy, T. Wald, G. Koehler, M. R. Rokuss, N. Disch, J. Holzschuh, D. Zimmerer, and K. H. Maier-Hein, "SAM.MD: Zero-shot medical image segmentation capabilities of the segment anything model," 2023, *arXiv:2304.05396*.
- [44] J. Lian, J. Liu, S. Zhang, K. Gao, X. Liu, D. Zhang, and Y. Yu, "A structure-aware relation network for thoracic diseases detection and segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 8, pp. 2042–2052, Aug. 2021, doi: [10.1109/TMI.2021.3070847](https://doi.org/10.1109/TMI.2021.3070847).
- [45] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: [10.1109/TPAMI.2018.2858826](https://doi.org/10.1109/TPAMI.2018.2858826).
- [46] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in *Machine Learning in Medical Imaging*. Cham, Switzerland: Springer, 2017, pp. 379–387, doi: [10.1007/978-3-319-67389-9\\_44](https://doi.org/10.1007/978-3-319-67389-9_44).
- [47] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Venice, Italy, Apr. 2019, pp. 683–687, doi: [10.1109/ISBI.2019.8759329](https://doi.org/10.1109/ISBI.2019.8759329).
- [48] S. Suzuki and K. Be, "Topological structural analysis of digitized binary images by border following," *Comput. Vis., Graph., Image Process.*, vol. 30, no. 1, pp. 32–46, Apr. 1985, doi: [10.1016/0734-189x\(85\)90016-7](https://doi.org/10.1016/0734-189x(85)90016-7).
- [49] M. F. Rahman, Y. Zhuang, T.-L. Tseng, M. Pokojovy, P. McCaffrey, E. Walsler, S. Moen, and A. Vo, "Improving lung region segmentation accuracy in chest X-ray images using a two-model deep learning ensemble approach," *J. Vis. Commun. Image Represent.*, vol. 85, May 2022, Art. no. 103521, doi: [10.1016/j.jvcir.2022.103521](https://doi.org/10.1016/j.jvcir.2022.103521).



field to advance healthcare technologies.

**EBRAHIM KHALILI** received the B.S. degree in control engineering from the K. N. Toosi University of Technology, Iran, in 2017, and the M.S. degree in biomedical engineering from Modares University, in 2020. Currently, he is a Research Assistant with the University of Cádiz, Spain, focusing on biomedical engineering in pulmonary diseases. His research interests include biomedical signals, image processing, brain-computer interfaces, and the dynamic neurocognitive exploration



**BLANCA PRIEGO-TORRES** is currently an Assistant Professor with the Department of Automation, Electronics Engineering, and a Computer Network Architecture with the School of Engineering, University of Cádiz, Spain. She is with the bioengineering, automation, and robotics research group. Her scientific research interests include artificial intelligence and image and signal processing.



**ANTONIO LEÓN-JIMÉNEZ** is currently a Doctor in medicine and leads the Pulmonology Unit, Puerta del Mar University Hospital, Cádiz. He is an Expert in COPD and interstitial lung diseases and leads the Andalusian Reference Unit for Silicosis caused by artificial stone. He has published more than 70 articles. He is a member of the Southern Pulmonology and Thoracic Surgery Association (Neumosur) and the Spanish Society of Respiratory Pathology (SEPAR).



**DANIEL SANCHEZ-MORILLO** is currently an Associate Professor with the Department of Automation, Electronics Engineering, and a Computer Network Architecture with the School of Engineering, University of Cádiz, Spain. He leads the bioengineering, automation, and robotics research group. His scientific research interests include signal and image analysis using artificial intelligence. He is a member of the Spanish Society of Biomedical Engineering (SEIB).

...