

A LIGHTWEIGHT U-NET WITH SEPARABLE CONVOLUTIONS FOR EFFICIENT LUNG SEGMENTATION IN REAL-TIME MEDICAL IMAGING

¹SALEH ALGHAMDI, ²MOHAMMAD SHORFUZZAMAN

¹Associate Professor, Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia

²Professor, Department of Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia

E-mail: s.algamedi@tu.edu.sa, msshapon2k@gmail.com

ABSTRACT

Accurate and efficient segmentation of the lung regions is indispensable for detecting and managing pulmonary diseases, as it allows clinicians to identify abnormalities and plan effective intervention strategies. However, the high computational demands of many existing segmentation models pose a significant challenge, particularly for deployment in resource-constrained environments such as mobile, edge platforms, and point-of-care devices. Lung segmentation is further challenged by wide anatomical variability and imaging artifacts, which existing models often struggle to handle without access to large-scale hardware. This study addresses this limitation by introducing a lightweight U-Net architecture that integrates depthwise separable convolutions to reduce computational complexity while preserving segmentation accuracy. By replacing standard convolutional layers, the model achieves faster inference and significantly lower parameter counts, making it well-suited for IT applications in embedded systems and clinical informatics. The model was evaluated on the publicly available *Pulmonary Chest X-Ray Defect Detection* dataset from Kaggle, demonstrating its effectiveness in segmenting lung regions. The performance evaluation shows that our model delivers outstanding results, attaining a Dice score of 91.92%, a Jaccard index of 82.75%, precision of 92.64%, recall of 90.31%, and accuracy of 97.12% on the test dataset. These results highlight that the lightweight U-Net achieves state-of-the-art segmentation accuracy with significantly reduced computational overhead, making it ideal IT solution for real-time use in clinical workflows and deployment on limited-resource devices.

Keywords: *Lung Segmentation, Depthwise Separable Convolutions, Lightweight U-Net, Chest X-ray Analysis, Real-Time Inference*

1. INTRODUCTION

This Lung cancer poses a major challenge to global health. In the United States, lung cancer ranks third in incidence but is the top cause of cancer-related deaths among both males and females. According to the American Cancer Society, approximately 226,650 new lung cancer cases are expected in the U.S. in 2025 (110,680 in males and 115,970 in females), with about 124,730 deaths (64,190 among males and 60,540 among females). Lung cancer represents about 11% of all new cancer diagnoses and accounts for 20% of cancer-related deaths [1]. According to the SEER database, in 2023, there were 226,650 new cases of lung cancer, and 124,730 deaths. The 5-year survival rate between 2015 and 2021 is 28.1% [2]. According to another study, respiratory illnesses such as bronchiolitis, bronchitis,

bronchopneumonia, interstitial pneumonia, lobar pneumonia, and pneumothorax rank among the leading causes of pediatric mortality in many countries [3].

Medical imaging is vital in contemporary healthcare, providing key insights for diagnosis, treatment planning, and monitoring disease progression. Chest radiography, in particular, is one of the most commonly performed radiological examinations due to its low cost, wide availability, and relatively low radiation dose [4]. Chest X-rays aid in diagnosing numerous lung disorders, such as pneumonia, tuberculosis, lung cancer, and other respiratory ailments. However, the manual analysis of chest X-ray images is time-consuming and subjective, leading to potential inter-observer variability and diagnostic delays. Computer-aided diagnosis (CAD) systems have been developed to

assist radiologists in interpreting medical images, improving the accuracy and efficiency of diagnosis. A fundamental step in many CAD systems is the accurate segmentation of the lung fields, which provides the anatomical context for further analysis [5]. It is an essential precursor to numerous downstream clinical tasks ranging from automated nodule detection and disease quantification to radiotherapy planning and longitudinal monitoring of pulmonary conditions. In resource-constrained or high-throughput settings such as emergency departments and portable screening stations, manual delineation of lung boundaries is time-consuming, subject to inter-observer variability, and often impractical. Thus, accurate and efficient lung region segmentation has become indispensable for effective diagnosis and treatment planning of pulmonary diseases. Effective lung segmentation enables physicians to identify abnormalities, quantify disease severity, and plan interventions with greater precision. However, the task presents several challenges due to the inherent variability in lung shapes and sizes, the presence of low contrast, pathological regions, and various imaging artifacts that can obscure lung boundaries [6].

Early traditional image-processing techniques such as global or adaptive thresholding, region growing, active contours, and edge detection offered computationally light solutions but repeatedly faltered when faced with low contrast, overlapping anatomy, or pathological abnormalities [7]. Classical machine-learning approaches (e.g., support vector machines or random forests operating on handcrafted features) improved robustness somewhat but remained limited by the representational capacity of hand-designed descriptors [8].

The advent of deep convolutional neural networks (CNNs) revolutionized medical image segmentation. U-Net, introduced by Ronneberger et al. [9], combined an encoder-decoder topology with skip-connections to simultaneously capture global context and fine spatial detail, rapidly becoming a de facto standard. Subsequent variants including SegNet [10], residual and dense U-Nets [11], attention-augmented architectures [12] and multiscale fusion models [13] have each pushed the state of the art in segmentation accuracy, achieving Dice scores often above 90 % on benchmark tasks. However, these gains come at the cost of ever-increasing model size (tens of millions of parameters) and longer inference times, which impede real-time deployment and use on edge or mobile devices.

Despite the advancements in medical image analysis, accurate and efficient lung segmentation remains a challenging problem. As seen, traditional image processing techniques often struggle with the inherent variability in lung shapes and sizes, the presence of low contrast, pathological regions, and various imaging artifacts that can obscure lung boundaries. This trade-off between segmentation reliability and computational efficiency constitutes our central research problem. How can we preserve the high accuracy of modern CNN-based methods while drastically reducing model size and latency? Depthwise separable convolutions, initially made popular in Xception [14], present a promising approach. They work by breaking down standard convolutions into more manageable depthwise and pointwise operations, which significantly reduces the number of parameters and FLOPs (floating point operations), while largely preserving the model's ability to accurately represent data. In this paper, we address this need by proposing a novel lightweight U-Net architecture that integrates separable convolutional neural networks to achieve efficient and accurate lung field segmentation in chest X-ray images. From an IT perspective, the significance of this work lies in its ability to bridge the gap between high-accuracy medical image segmentation and practical deployment on real-world IT systems. Traditional deep learning models often assume access to extensive computational resources, which is unrealistic for embedded healthcare platforms, mobile diagnostics, and low-power edge devices. By designing a compact, fast, and accurate segmentation model, we contribute to the growing field of intelligent medical imaging within IT infrastructure, supporting scalable and cost-effective solutions for clinical workflows. The model's lightweight architecture enables integration into hospital PACS (picture-archiving and communications systems), mobile applications, and AI-enabled radiology tools, advancing the application of IT in digital health ecosystems.

In summary, this research endeavors to address key questions surrounding the potential of a lightweight U-Net architecture, specifically investigating whether it can match or exceed the segmentation accuracy of a standard U-Net on a large chest X-ray dataset, while also evaluating its computational and memory footprint, including the feasibility of achieving sub-20 millisecond inference on a high-end GPU. The study further aims to compare the model's performance against classical, CNN-based, and GAN-based segmentation baselines, and to explore the potential clinical applications of accurate and efficient lung

segmentation in enhancing patient care and outcomes. The key contributions of this research are summarized as follows:

a) We introduce a 2D U-Net variant employing depth-wise separable convolutions in every layer, reducing parameter count and FLOPs by over an order of magnitude without sacrificing segmentation quality,

b) We demonstrate sub-20 millisecond inference on a Tesla P100 GPU for 512×512 chest X-rays enabling practical integration into time-sensitive clinical workflows and resource-limited devices,

c) We provide a comprehensive evaluation of the model's performance, including accuracy, Dice coefficient, Jaccard index, precision, and recall, highlighting its superior segmentation performance with reduced computational overhead,

d) We conduct a comprehensive quantitative and qualitative analysis on a benchmark dataset [15], comparing against both classical and deep-learning baselines, and provide ablation studies to isolate the impact of separable convolutions,

e) We discuss the potential for deployment on edge hardware and integration into downstream tasks illustrating the model's utility in real-world settings.

The rest of this paper is structured as follows. Section 2 surveys existing methods for lung segmentation. Section 3 details the proposed lightweight U-Net architecture and the experimental setup. Section 4 presents experimental results and comparative analysis. Section 5 discusses the conclusions and future work.

2. LITERATURE REVIEW

Conventional lung segmentation approaches initially made extensive use of image-processing methods such as thresholding, edge detection, region growing, and active contours [7]. While these methods are computationally efficient, they often struggle with anatomical variability, low contrast, pathological regions, and imaging artifacts. To overcome these limitations, machine learning techniques, including support vector machines and random forests, were introduced, offering modest improvements through handcrafted features [8]. Lately, deep learning, particularly convolutional neural networks (CNNs), has showed amazing success in segmenting medical images, often surpassing traditional methods in terms of accuracy and robustness.

Classical methods for lung segmentation in medical images, particularly Computed Tomography (CT) and X-ray, often rely on image processing techniques based on intensity, shape, and anatomical knowledge. These methods typically involve a pipeline of steps, including preprocessing, segmentation, and postprocessing. While effective in many cases, they can struggle with pathological lungs, low contrast images, and variations in lung shape and size [7].

Segmenting lungs with Juxta-Pleural nodules is a complex task where traditional methods like thresholding, region-growing, and active contours often fail. To address this, a fully automated method is proposed in [16] with two stages: lung field extraction and boundary analysis for accurate segmentation of lungs including Juxta-Pleural nodules. The proposed method outperforms traditional thresholding techniques. Moreover, traditional machine learning methods for lung segmentation often involve manual feature extraction based on grayscale, geometric shapes, or anatomical knowledge. These methods are generally less effective in handling complex variations and pathologies in lung images [17]. Additionally, traditional approaches are typically more time-consuming and less adaptable to new data compared to deep learning methods, which can automatically learn and adapt to new patterns [18].

In recent years, deep learning, especially Convolutional Neural Networks (CNNs), has revolutionized medical image segmentation, including lung segmentation. These methods learn intricate features directly from the image data, leading to more robust and accurate segmentation, particularly in challenging cases. The U-Net model, first presented by Ronneberger et al. [9], has evolved into a key framework for segmenting medical images. The encoder-decoder design of U-Net, augmented by skip connections, facilitates accurate localization and allows effective training even on small datasets. Numerous studies have adapted and improved the U-Net for lung segmentation. For instance, Badrinarayanan et al. [10] introduced SegNet, a deep encoder-decoder architecture with max-pooling indices used in upsampling, achieving robust results in semantic segmentation tasks including lung CTs. Hofmanninger et al. [4] applied a U-Net variant to segment lungs from chest CTs, demonstrating high performance across multiple datasets and emphasizing the importance of data diversity. Similarly, Hwang et al. [19] used U-Net for lung field segmentation on chest X-rays and addressed

challenges such as overlapping clavicles and heart regions. To better capture spatial context and reduce false positives, multiscale architectures have been explored. Tang et al. [20] proposed a multiscale feature fusion U-Net for segmenting infected regions in COVID-19 CT images, effectively capturing both coarse and fine structures. The integration of residual connections and dense blocks has also improved feature propagation and gradient flow. Jin et al. [21] presented a residual attention U-Net to segment lung tumors from CT volumes, achieving superior boundary delineation.

Gite et al. [22] discussed the implementation of U-Net++ for lung segmentation using X-ray images. The model achieves over 98% segmentation accuracy and a mean intersection over union of 0.95, demonstrating its efficacy in diagnosing pulmonary diseases. However, the lack of practical application can raise questions about how well the models would perform in actual clinical settings. The TVAC (Total Variation-based Active Contour) algorithm is proposed in [6] for accurate lung segmentation in chest X-rays, particularly in critically ill patients. This method is particularly effective in handling the complexities of lung segmentation in the presence of medical equipment and varying patient conditions. It shows moderate performance, achieving a Dice coefficient of 0.86 for adults and 0.85 for children. Khomduean et al. [23] developed a model combining 3D-UNet with DenseNet169 and ResNet to segment lung lobes and lesions, achieving Dice similarity coefficients of 91.52% and 76.89%, respectively. Delfan et al. [24] introduced CT-LungNet, a fully automatic method for segmenting lung tissue in 3D CT images. The model employs a 2.5D image representation and a U-Net architecture with pre-trained InceptionV3 blocks, aiming to reduce the number of learnable parameters while maintaining high segmentation accuracy. Evaluated on public datasets LUNA16, VESSEL12, and CRPF, CT-LungNet achieved Dice coefficients of 99.7%, 99.1%, and 98.8%, respectively. While CT-LungNet demonstrated high performance on public datasets, its generalizability to diverse clinical settings with varying imaging conditions remains to be validated. Also, the dependency on pre-trained architectures may limit their effectiveness in specific application areas. In a subsequent study, Wu et al. [25] characterizes a considerable advancement in lung nodule semantic segmentation, addressing challenges such as under and over segmentation in CT images. This model enhances the traditional U-Net architecture by integrating a ResNet encoder, atrous spatial pyramid pooling,

and a cross-fusion feature module with attention mechanisms, leading to improved segmentation accuracy. The model achieved a mean Intersection over Union (mIoU) of 87.76% and an F1-score of 93.56% on the LIDC dataset, outperforming existing models like SegNet and U-Net. The model's effectiveness in diverse clinical scenarios and its integration into existing diagnostic workflows need further exploration.

While RAD-UNet shows promising results, other models like improved V-Net [12] and context-aware attention U-Net [5] also demonstrate competitive performance, suggesting a diverse landscape of effective segmentation techniques in lung nodule detection. The V-Net model combines pixel threshold segmentation with an attention mechanism, aiming to enhance the segmentation process for lung nodules. It achieves high Dice similarity coefficients and sensitivity on public datasets LUNA16 and LNDb. On the other hand, the complementary context-aware (CCA) attention module in [5] focuses on a coarse-to-fine 3D segmentation framework for lung nodule segmentation in CT images. Designed to enhance segmentation accuracy, the CCA module effectively captures 3D spatial dependencies and complex contextual information. A recent study by Kongkham et al. [26] compares deep learning methods with traditional techniques using two widely used datasets, likely LIDC-IDRI and LUNA16. The main evaluation metric is the Sorensen-Dice Coefficient (DSC), which measures how well the segmentation matches the ground truth. Results show that deep learning significantly outperforms traditional methods achieving a DSC of 0.853 vs. 0.761 on the first dataset, and 0.763 vs. 0.704 on the second. These findings highlight deep learning's strength in handling the complexity of lung nodule segmentation.

Din et al. [27] introduced CXR-Seg, a deep learning model designed for precise lung segmentation in chest X-ray images. The architecture integrates a pre-trained EfficientNetV2S encoder with a spatial enhancement module (SEM), transformer attention module (TAM), and multi-scale feature fusion block (MS-FFB) to enhance feature representation and capture contextual information. Evaluated on four public datasets such as Montgomery, Shenzhen, Darwin COVID-19, and TCIA, the model achieved high performance metrics, including a Dice coefficient of up to 97.76%. While CXR-Seg demonstrated strong performance on public datasets, its generalizability to diverse

clinical settings with varying imaging conditions remains to be validated. Additionally, the computational complexity introduced by the integrated modules may pose challenges for deployment in resource-constrained environments. In another effort, In [11], Alam et al. presented AMRU++, a UNet++ variant with multi-residual blocks and an attention mechanism, optimized for segmenting lungs in chest radiographs especially those showing advanced conditions such as pneumoconiosis, COVID-19, and tuberculosis. The model incorporates attention modules and multi-residual blocks to capture relevant spatial information and rich contextual features. To address the scarcity of annotated pathological data, a novel data augmentation technique was introduced, simulating disease-specific features to enhance model robustness. With a Dice score of 0.9363, AMRU++ outperformed several state-of-the-art architecture. Despite its high performance, AMRU++'s reliance on complex architectural components and extensive data augmentation may increase computational demands. Moreover, the effectiveness of the augmentation technique in representing the full spectrum of pathological variations requires further investigation.

Cai et al. [28] proposed a novel lung segmentation approach using generative adversarial networks (GANs), specifically leveraging the Pix2Pix framework for image-to-image translation tasks. In their method, the original CT lung images are treated as “blurred” inputs and the ground truth segmentation maps as “clear” outputs. The GAN learns to translate the input images into accurate segmentation masks by minimizing both pixel-wise loss and adversarial loss. Despite its promising results, the study is limited by its reliance on a relatively small dataset, which may affect generalizability. Additionally, the method focuses on 2D slice-based segmentation and does not yet address the full 3D volumetric nature of CT scans, leaving room for future improvements in clinical applications. Multimodal fusion techniques proposed in [29] involve combining features from CT and PET images at different stages of the U-Net architecture. In particular, the model incorporates various fusion strategies, including hyper-dense fusion and hyper-dense VGG16 U-Net, to integrate anatomical and functional data. Some other studies [30-34] also developed multi-modal lung tumor segmentation network combining CT and PET images. In a recent study, Goswami et al. [35] presented a lung segmentation method designed to work effectively even with limited medical imaging data, a common issue due to legal and privacy

constraints. Unlike traditional U-Net models that need large datasets and long training times, the proposed method combines a U-Net-style generator with a generative adversarial network (GAN). The generator creates synthetic images to help balance the dataset and better represent rare cases, focusing specifically on lung regions. It uses an encoder-decoder structure with skip connections to retain important features, while the discriminator helps the generator improve by distinguishing real from fake images. The model achieved an accuracy of 84.39%, showing strong performance even with restricted training data. Zafaranchi et al. [36] proposed a deep learning-based framework for lung nodule detection and segmentation using the LUNA-16 dataset. The method follows a two-phase pipeline: lung segmentation using the LungQuant algorithm, followed by nodule segmentation via a fine-tuned Attention Res-UNet. The system achieved an average Dice Similarity Coefficient (DSC) of 90% for lung segmentation and 81% for nodule segmentation, indicating high accuracy. To enhance model interpretability, Grad-CAM was applied, supporting its clinical relevance. However, the model's evaluation was limited to a single dataset, which may restrict its generalizability; further testing on diverse datasets and in real-world clinical settings is needed to confirm its robustness.

While advancements in deep learning have led to highly accurate lung segmentation models, many existing architectures, particularly standard U-Net and its variants, remain computationally intensive and dependent on large, annotated datasets and prolonged training. These limitations hinder their adoption in real-time and resource-constrained clinical settings. To address this, our study introduces a streamlined U-Net model incorporating depth-wise separable convolutions, which significantly reduces the number of parameters and training time while maintaining high segmentation accuracy and enabling faster inference. This approach bridges the gap between state-of-the-art performance and practical usability, making it well-suited for real-world deployment in diverse healthcare environments.

3. MATERIALS AND METHODS

This section outlines the workflow depicted in Figure 1 for performing accurate and efficient lung segmentation on chest X-rays, optimized for real-time execution. Our approach centers around a lightweight yet powerful 2D U-Net architecture, specifically designed to minimize computational overhead while preserving crucial segmentation

capabilities for this modality. To enhance the network's efficiency, we employ depth-wise separable convolutions throughout the architecture, significantly reducing the number of parameters and computations suitable for potentially resource-constrained environments. The complete workflow involves preprocessing the input chest X-ray images, feeding them into our lightweight 2D U-Net built with separable convolutions, and finally, postprocessing the resulting segmentation masks to refine the output. More specifically, the preprocessed images pass through the encoder path, which captures contextual information and extracts

separable convolutions within the U-Net, and the experimental setup using chest X-ray datasets. The complete step-by-step segmentation process is also outlined in Algorithm 1.

3.1 Data Acquisition and Preprocessing

In this study, we employed the Chest X-ray Masks and Labels dataset [33], publicly available on Kaggle, to train, validate, and evaluate the proposed lung segmentation model. The dataset consists of paired grayscale chest radiographs typically captured in posterior-anterior (PA) or

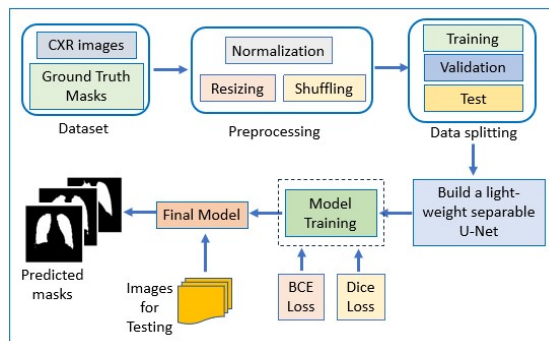


Figure 1. Workflow of the proposed lightweight U-Net with separable convolutions for real-time lung segmentation on mobile, edge, and point-of-care platforms.

hierarchical features using depth-wise separable convolutions. The subsequent bottleneck layer represents the network's deepest stage and provides the connection between encoder and decoder. The decoder path then progressively upsamples the feature maps to reconstruct spatial dimensions. The last convolutional layer then reduces the channel depth to match the number of target classes, producing the final segmentation map. Finally, the postprocessing step refines the raw segmentation output to produce the final lung mask using binary thresholding which converts the probabilistic map into a binary mask by applying a threshold. The final output of the workflow is a binary mask accurately delineating the lung regions in the input image. The model is trained using the binary cross-entropy loss function to optimize performance, and the best-performing version is selected for predicting the segmentation masks of the test images. The workflow illustrates an efficient and structured approach for segmenting lung regions from chest radiographs with the proposed lightweight U-Net. The subsequent subsections will elaborate on each component of this workflow, including the network architecture tailored for 2D X-ray data, the implementation details of the

Algorithm 1: Lightweight U-Net with depthwise separable convolutions for fast and accurate lung segmentation

Input: Chest radiographs and their associated ground truth segmentation labels

Preprocess the input images by resizing to 256×256 pixels and scaling their intensities in the range $[0, 1]$

Divide the data into training and test subsets

Construct a lightweight 2D U-Net architecture using depth-wise separable convolutions in both encoder and decoder paths

Train the network with the Adam optimizer and binary cross-entropy as the loss metric

Predict segmentation masks for the test set using the trained model

Postprocess the predicted masks using binary thresholding to convert the probabilistic output into a binary lung mask

Output: Final binary masks accurately delineating lung regions in CXr images

anterior-posterior (AP) views and their corresponding binary segmentation masks that delineate the left and right lung regions. Each mask highlights the lung areas using a distinct pixel intensity (255) against a black background (0), making them well-suited for binary segmentation tasks. Figure 2 depicts sample chest radiographs and their corresponding ground truth masks used for lung segmentation.

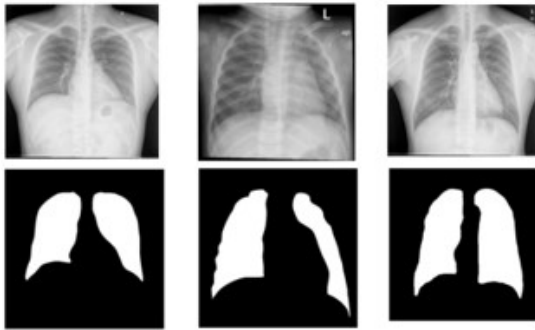


Figure 2. Example chest X-ray images with corresponding lung masks

This dataset is specifically curated for supervised learning in medical image segmentation and serves as a widely accepted benchmark for developing and assessing automated lung segmentation algorithms. It plays a crucial role in enabling models to learn accurate lung localization, which is often the first step in computer-aided diagnosis and quantitative pulmonary assessment from chest X-rays. Although the dataset includes a large number of image-mask pairs, some samples may be missing or misaligned, necessitating manual verification and filtering to ensure data integrity. The dataset initially includes 800 chest X-ray images and 703 corresponding binary masks. Since 97 masks were missing, only the 703 CXR images with valid masks were used to maintain consistency and ensure reliable supervised training. The images and masks are provided in standard .png format, allowing seamless integration with deep learning pipelines. Although the dataset aggregates samples from open-access medical repositories and research datasets, it provides a substantial volume and diversity to support training of robust and generalizable deep learning models for lung segmentation. Manual verification and filtering were applied during preprocessing to ensure data integrity and alignment between image-mask pairs.

In this study, the chosen CXR dataset underwent a series of preprocessing steps to ensure optimal model training and evaluation. Initially, the dataset was partitioned into training and testing subsets, with 80% of the images allocated for training the model and the remaining 20% reserved for evaluating its performance on unseen data. This split allows for an unbiased assessment of the model's generalization capability.

Subsequently, each chest X-ray image was resized to a uniform spatial dimension of 512 x 512 pixels. This resizing operation serves multiple crucial purposes. Firstly, it standardizes the input size for the neural network, ensuring that all images have consistent dimensions regardless of their

original acquisition parameters. This uniformity is essential for efficient batch processing and stable gradient calculations during training. Secondly, resizing can help to reduce computational complexity, especially if the original images have very high resolutions, without significantly sacrificing the essential anatomical information required for lung segmentation.

Following resizing, the pixel intensity values of all images were rescaled to the range of 0 to 1. Medical images often have pixel intensities spanning a wide range. Normalizing these values to a smaller, consistent range like [0, 1] offers several benefits for deep learning models. It helps to prevent large intensity values from dominating the learning process, leading to more stable and faster convergence during training. Furthermore, it ensures that all input features are within a similar scale, which is generally beneficial for the performance of gradient-based optimization algorithms.

Finally, the training dataset was shuffled randomly before being fed into the model during each epoch. Shuffling is a standard practice in machine learning that helps to break any potential order-related biases within the dataset. By presenting the model with a different order of training samples in each epoch, we aim to improve the model's ability to learn generalizable features from the entire training distribution and prevent it from memorizing the order of the data, thus contributing to a more robust and less overfit model.

3.2 Proposed Model Architecture

Our proposed efficient lung segmentation framework leverages a 2D U-Net architecture [9] as its foundational backbone. The U-Net has demonstrated remarkable efficacy in various biomedical image segmentation tasks due to its characteristic encoder-decoder structure coupled with skip connections. This architecture is specifically well-suited for tasks where precise localization and contextual understanding are crucial. In our implementation, we aim for real-time performance by significantly reducing the model's computational footprint through the strategic replacement of standard convolutional layers with depth-wise separable convolutions.

3.2.1 Enhanced U-Net Variant

The modified U-Net architecture as shown in Figure 3 comprises two main paths: a contracting path (encoder) and an expansive path (decoder).

Contracting Path (Encoder): The encoder path follows a convolutional neural network structure. It

consists of a series of repeated blocks, where each block contains a pair of 3×3 depth-wise separable convolution layers, with each layer immediately followed by batch normalization and a ReLU activation. The number in the parenthesis after the kernel size indicates the number of output channels for that block. After each block, a 2×2 max-pooling operation with a stride of 2 is applied for downsampling the feature maps, effectively doubling the number of feature channels while halving the spatial dimensions. This process progressively extracts hierarchical features from the input image, capturing both local details and global context.

Mathematically, let $X_i \in \mathbb{R}^{H_i \times W_i \times C_i}$ be the input to the i -th encoder block. The block performs two convolutional operations, denoted by Conv_1

and Conv_2 , followed by BN and ReLU activation (σ):

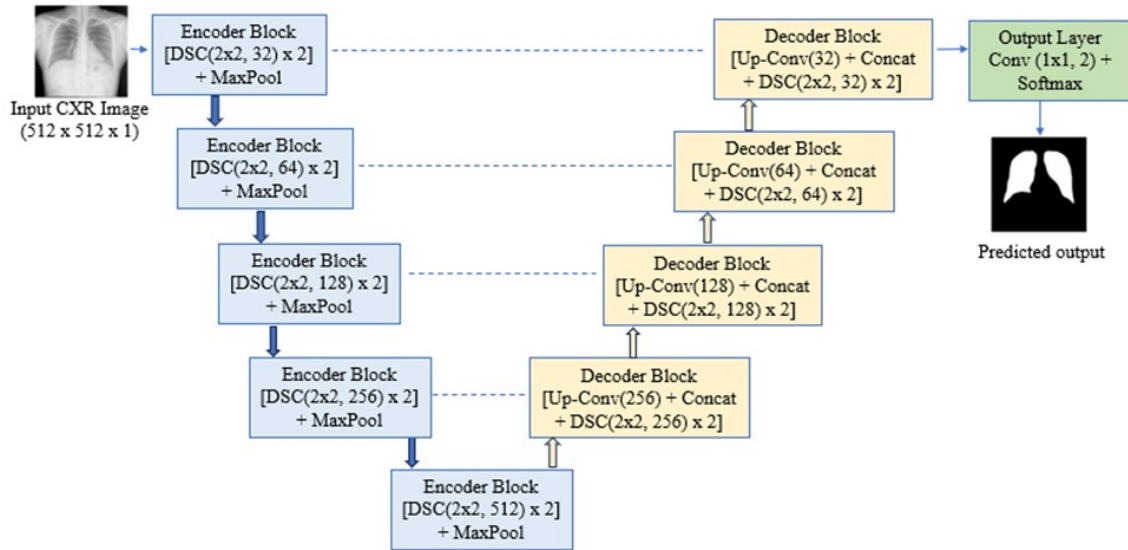
$$F_{i,1} = \sigma(\text{BN}(\text{Conv}_1(X_i))) \quad (1)$$

$$F_{i,2} = \sigma(\text{BN}(\text{Conv}_2(F_{i,1}))) \quad (2)$$

The output of the i -th encoder block, $E_i = F_{i,2}$, is then passed to a max-pooling layer MaxPool for downsampling:

$$X_{i+1} = \text{MaxPool}(E_i) \quad (3)$$

This downsampling continues until the bottleneck layer is reached, where the feature maps have the smallest spatial dimensions and the largest number of channels, representing the most abstract features of the input image.



Legend

DSC($k \times k, C$): Two consecutive depthwise separable convolution layers with a kernel size of $k \times k$ and C output channels for the second DSC layer.

Concat: Channel-wise concatenation with the corresponding encoder feature map.

Up-Conv(C): Transposed convolution for upsampling to C channels.

Figure 3. Block diagram of the proposed lightweight U-Net architecture with separable convolutions.

Expansive Path (Decoder): The decoder path mirrors the encoder path in reverse, gradually reconstructing the spatial resolution of the input image to produce a segmentation map. It consists of a sequence of upsampling operations, typically implemented using transposed convolutions (deconvolutions), each of which doubles the spatial dimensions of the feature map. After each upsampling step, the resulting feature map is concatenated channel-wise with the corresponding feature map from the encoder via skip connections.

This fusion provides high-resolution spatial details from the encoder, enabling the decoder to better localize and refine segmented regions.

Following the concatenation, the combined feature maps are passed through two consecutive DSC layers, each followed by BN and ReLU activation. These DSC layers integrate information from both the upsampled decoder features and the encoder's high-resolution features while reducing the number of parameters and computational cost. The number of output channels after concatenation

is the sum of the channels from both sources; the DSC layers then reduce and refine these channels to prepare for the next decoding stage. This process is repeated until the final output feature map matches the input image's spatial resolution, with the number of output channels corresponding to the number of segmentation classes (lung and background).

Mathematically, let $D_j \in R^{H_j \times W_j \times C_j}$ be the input to the j -th decoder block. First, it undergoes a transposed convolution ($Conv^T$) for upsampling:

$$U_j = Conv^T(D_j) \quad (4)$$

This upsampled feature map U_j is then concatenated (Concat) with the corresponding feature map E_{n-j} from the encoder (where n is the total number of downsampling steps):

$$M_j = Concat(U_j, E_{n-j}) \quad (5)$$

Subsequently, two convolutional operations, each followed by BN and ReLU, are applied:

$$F'_{j,1} = \sigma(BN(Conv_1(M_j))) \quad (6)$$

$$F'_{j,2} = \sigma(BN(Conv_2(F'_{j,1}))) \quad (7)$$

The output of the j -th decoder block is $D_{j+1} = F'_{j,2}$. This upsampling and convolution process continues until the final layer.

Skip Connections: The skip connections are a crucial aspect of the U-Net architecture. They directly connect the feature maps from the encoder path to the decoder path at the corresponding resolution levels. This allows the decoder to recover fine-grained details lost during the downsampling process in the encoder, leading to more accurate and spatially precise segmentations.

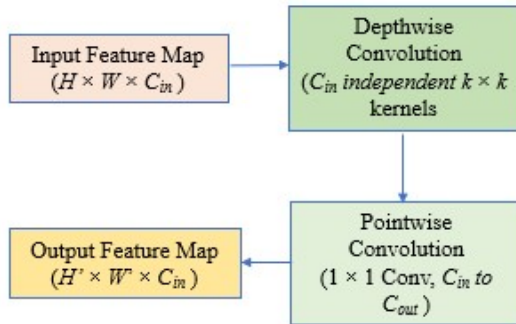


Figure 4. Illustration of the depthwise separable convolution (DSC) operation

3.2.2 Depthwise Separable Convolutions

To achieve a lightweight and efficient model suitable for real-time lung segmentation, we replace all standard convolutional layers within the U-Net architecture with depthwise separable convolutions as shown in Figure 4. This factorization of the standard convolution operation significantly

achieves a substantial reduction in both parameter count and computational workload.

A standard 2D convolution operates on all channels of the input feature map simultaneously to produce each channel of the output feature map. In contrast, a depthwise separable convolution breaks this process into two distinct steps: depthwise convolution and pointwise convolution.

Depthwise Convolution: The depthwise convolution applies a single convolutional filter to each input channel independently. Given an input feature map $X_i \in R^{H \times W \times C_{in}}$ a set of C_{in} number of $k \times k$ depthwise kernels $K_{dw} \in R^{k \times k \times 1 \times C_{in}}$ (where each kernel of size $k \times k \times 1$ operates on a single input channel), produces an intermediate feature map $Z \in R^{H \times W \times C_{in}}$. The element at position (i, j, c) in the output channel c of the intermediate map is computed by convolving the c -th channel of the input feature map with the c -th depthwise kernel:

$$Z(i, j, c) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} X(i+m, j+n, c) \cdot K_{dw}(m, n, 0, c) \quad (8)$$

The number of parameters in the depthwise convolution layer is $k \times k \times C_{in}$.

Pointwise Convolution: The pointwise convolution is a standard 1×1 convolution applied to the output of the depthwise convolution. It linearly combines the channel outputs of the depthwise convolution to create new features. Given the intermediate feature map $Z \in R^{H \times W \times C_{in}}$ and a set of C_{out} number of 1×1 pointwise kernels $K_{pw} \in R^{1 \times 1 \times C_{in} \times C_{out}}$ the final output feature map $Y \in R^{H \times W \times C_{out}}$ is calculated as:

$$Y(i, j, c_{out}) = \sum_{c_{in}=0}^{C_{in}-1} Z(i, j, c_{in}) \cdot K_{pw}(0, 0, c_{in}, c_{out}) \quad (9)$$

The number of parameters in the pointwise convolution layer is $1 \times 1 \times C_{in} \times C_{out} = C_{in} \times C_{out}$.

Parameter Efficiency: The total number of parameters in a depthwise separable convolution block is the sum of the parameters in the depthwise and pointwise stages:

$$N_{separable} = N_{sep} = (k^2 \times C_{in}) + (C_{in} \times C_{out}) \quad (10)$$

The number of parameters in a standard convolution with the same input and output channel dimensions and kernel size is:

$$N_{standard} = N_{std} = (k^2 \times C_{in} \times C_{out}) \quad (11)$$

The ratio of parameters in a depthwise separable convolution to a standard convolution is:

$$N_{sep} / N_{std} = ((k^2 \times C_{in}) + (C_{in} \times C_{out})) / (k^2 \times C_{in} \times C_{out}) = 1/C_{out} + 1/k^2 \quad (12)$$

For typical CNN architectures where the number of output channels (*Cout*) and the kernel size (*k*) are significantly larger than 1, the use of depthwise separable convolutions leads to a substantial reduction in the number of trainable parameters and the associated computational cost, making the network more efficient and potentially suitable for real-time applications with limited computational resources. By replacing the standard convolutions in the U-Net with these efficient depthwise separable convolutions, our proposed model aims to achieve a favorable trade-off between segmentation accuracy and computational efficiency for lung segmentation in chest X-ray images.

3.3 Experimental Setup and Evaluation

This section outlines the experimental protocol employed to evaluate the performance of the proposed lightweight U-Net architecture with depthwise separable convolutions for lung segmentation in chest X-ray images. It details the implementation specifics, training parameters, and evaluation metrics.

3.3.1 Implementation Details

The implementation and experimentation of the proposed lightweight U-Net architecture were conducted using Python version 3.8, leveraging the deep learning framework TensorFlow version 2.6. All model training and evaluation processes were accelerated using the computational resources provided by Kaggle, specifically utilizing an NVIDIA Tesla P100 Graphics Processing Unit (GPU). This GPU acceleration significantly reduced the training time required for the deep learning model, enabling efficient experimentation and validation within the Kaggle environment.

Network Configuration: As part of the network configuration, encoder path consists of four sequential EncoderBlock layers with filter sizes of 32, 64, 128, and 256, respectively. Each EncoderBlock utilizes two consecutive SeparableConv2D layers with a kernel size of 2x2, ReLU activation, and 'same' padding, followed by a Dropout layer with varying dropout rates (0.1 for the first two encoder blocks and 0.2 for the subsequent ones). Max-pooling with a default pool size of 2x2 is applied after each EncoderBlock to downsample the feature maps, except for the final encoding layer. The bottleneck of the U-Net consists of two SeparableConv2D layers with 512 filters, a 2x2 kernel size, ReLU activation, 'same' padding, and a dropout rate of 0.3, without any subsequent pooling. The decoder path mirrors the encoder with four DecoderBlock layers, having

filter sizes of 256, 128, 64, and 32. Each DecoderBlock upsamples the incoming feature map using UpSampling2D and concatenates it with the corresponding skip connection from the encoder path. This concatenated feature map is then processed by an EncoderBlock (without pooling) with the specified filter size, kernel size of 2x2, and dropout rate (0.2 for the first two decoder blocks and 0.1 for the subsequent ones). Finally, a 1x1 Conv2D layer with a sigmoid activation function and 'same' padding is used to create the final segmentation output with a single channel representing the probability of each pixel belonging to the lung class.

Loss Function and Optimization: The model was trained using the binary cross-entropy (BCE) loss function. BCE is a standard loss function for binary classification tasks, such as segmenting the lungs (positive class) from the background (negative class). This metric assesses the pixel-level discrepancy between the predicted probability outputs and the actual binary labels. Mathematically, for a single pixel i with true label $y_i \in \{0,1\}$ and predicted probability $p_i \in [0,1]$, the binary cross-entropy loss L_{BCE} is calculated as:

$$L_{BCE}(y_i, p_i) = -(y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (13)$$

While the primary loss function used for optimization was binary cross-entropy, the Dice coefficient loss (*dice_coef_loss*) was also used as a metric during training. Because it's based on the Dice similarity coefficient, the Dice loss directly reflects the overlap between predicted and actual segmentation masks. Minimizing the Dice loss effectively maximizes the Dice coefficient, providing a training signal that is directly aligned with the desired segmentation performance.

The network was optimized with the Adam algorithm using its default learning rate settings in TensorFlow 2.6. As an adaptive optimizer, Adam computes separate learning rates for each model parameter, often leading to faster convergence and good performance. The model was trained with batches of 16 samples. The batch size specifies how many samples the model processes before each weight update. A batch size of 16 represents a trade-off between computational efficiency and reliable gradient estimation. Training was conducted for 100 epochs in total. The epoch count was determined to provide sufficient training iterations for the model to discover key patterns and converge on optimal weights for lung segmentation. The training progress was monitored using the defined loss function and evaluation metrics, potentially with

early stopping to prevent overfitting and optimize the training duration.

3.3.2 Assessment Criteria

The performance of the proposed lung segmentation model was quantitatively evaluated using a comprehensive set of metrics, including the dice similarity coefficient (DSC), intersection over union (IoU), accuracy, precision, and recall. These metrics were chosen to provide a comprehensive assessment of the model's ability to accurately delineate the lung regions in chest X-ray images. The DSC measures how much the predicted segmentation overlaps with the ground truth mask. It is especially useful in medical imaging tasks where class imbalance is common, as it gives more importance to correctly segmented regions. A higher dice score indicates better agreement between the predicted and actual lung regions. IoU also known as the Jaccard Index, measures the proportion of overlap between the predicted and true segmentation regions relative to their total combined area. It is a strict measure of segmentation accuracy and is useful for comparing performance across different models. Again, a higher IoU is better. Accuracy represents the fraction of pixels correctly labeled as either lung or background. While it provides an overall sense of model performance, it can be misleading in imbalanced datasets where background pixels dominate. Precision indicates how many of the predicted lung pixels actually belong to the lung region, while recall reflects how many of the true lung pixels were correctly identified by the model. These metrics are especially important when assessing the clinical reliability of the segmentation model, ensuring it avoids both over-segmentation (false positives) and under-segmentation (false negatives). Together, these metrics offer a robust and well-rounded evaluation of the model's

segmentation performance, supporting both technical assessment and potential clinical applicability.

3.3.3 Evaluation Procedure

The trained lightweight U-Net model was evaluated on the held-out test dataset, which comprised 20% of the initially split data and contained images the model had never seen during training. For each test image, the model produced a probability map showing how likely each pixel belongs to the lung region. We then binarized this probability map into a segmentation mask by applying a fixed threshold of 0.5.

After producing binary masks for all test images, we computed the evaluation metrics by comparing these predictions to the dataset's ground truth

masks. These metrics were computed on a per-image basis, and the reported results represent the average of these metrics across the entire test set. This procedure yields quantitative metrics that reflect the model's generalization to new data and its precision in segmenting lung regions on chest X-rays.

4. RESULTS ANALYSIS

This section presents a comprehensive analysis of the lung segmentation performance achieved by the proposed lightweight U-Net architecture with depthwise separable convolutions. The quantitative evaluation of our model was conducted using key metrics including the DSC, IoU, accuracy, precision, and recall, calculated on the held-out test dataset. To provide a visual understanding of the segmentation quality, we present qualitative comparisons between the predicted segmentation masks and the corresponding ground truth annotations. Furthermore, we illustrate the training progress of our model through a qualitative visualization of predicted masks on a sample test image at various training epochs. Furthermore, the proposed model's performance was benchmarked against results reported in existing literature to assess its relative strength and generalizability.

4.1 Evaluation of Lung Segmentation Performance

In this section, we report the quantitative evaluation results of our lightweight U-Net on the chest X-ray dataset. We assessed the model's segmentation capabilities using standard performance metrics, across training, validation, and held-out test datasets. Table 1 provides a summary of the results for each dataset.

Table 1. Quantitative performance of the proposed segmentation model

Data	Performance Metric				
	Acc (%)	Dice (%)	IoU (%)	Prec (%)	Rec (%)
Training	98.89	93.88	86.44	94.09	93.38
Validation	97.78	92.70	81.64	90.79	88.38
Test	97.12	91.92	82.75	92.64	90.31

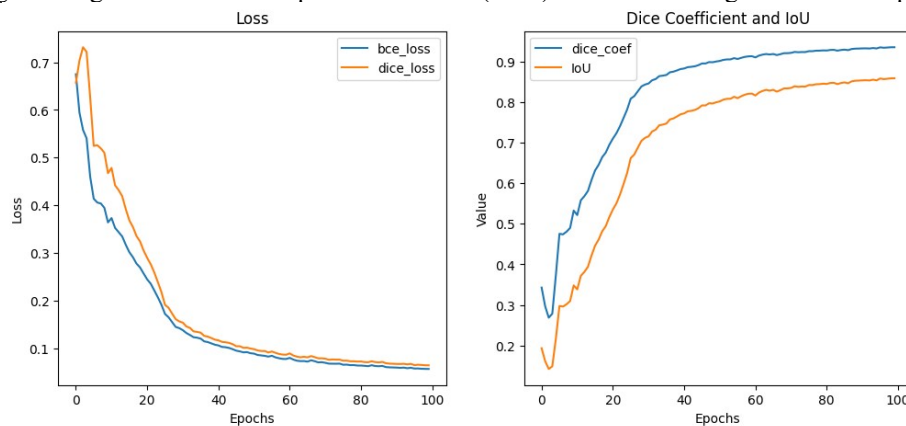
The proposed model achieved a strong performance on the unseen test dataset. The accuracy of 97.12% indicates a high pixel-level classification rate, with the model correctly identifying the majority of pixels as either lung or

background. With a Dice coefficient of 91.92%, the predicted lung segmentation closely matches the ground truth annotations, highlighting precise volumetric segmentation. Similarly, the IoU of 82.75% further corroborates the significant spatial agreement between the predicted and actual lung masks. In terms of error types, the precision of 92.64% suggests that when the model predicted a pixel as belonging to the lung, it was correct a high percentage of the time, indicating a low rate of false positives. On the other hand, the recall of 90.31% indicates that the model successfully identified a large proportion of the actual lung pixels, suggesting a low rate of false negatives.

Overall, the performance metrics were generally consistent across the training, validation, and test datasets, suggesting good generalization of the model. A slightly higher performance on the training set is expected as the model has been directly optimized on this data. The validation set performance, which is used for hyperparameter tuning and early stopping, shows a minor decrease compared to the training set, indicating a good balance between learning and generalization. Although performance on the test set is marginally lower than on the validation set, it still confirms the proposed lightweight U-Net's robustness and effectiveness for lung segmentation. These results confirm that the proposed lightweight U-Net model with depthwise separable convolutions is capable of achieving high segmentation accuracy while maintaining a strong balance between precision and

recall across different data splits. The consistently high Dice and IoU scores across all datasets highlight the model's ability to accurately capture the shape and location of the lungs.

Figure 5(a) illustrates the training loss curves for Binary Cross-Entropy (BCE) loss and Dice loss over 100 training epochs. Both loss functions show a consistent downward trend, indicating successful convergence of the model during training. During the initial training phase (epochs 0–20), both loss curves drop sharply, indicating the model's rapid acquisition of basic lung region features. The BCE loss starts slightly lower and decreases more steadily compared to the Dice loss, which begins higher but catches up as training progresses. This behavior is expected since Dice loss is more sensitive to class imbalance and segmentation overlap, often taking longer to stabilize in the early training stages. From epoch 20 onward, both loss curves flatten gradually, showing continued but slower improvement. By epoch 60 and beyond, the losses reach near-convergence, with values stabilizing below 0.1, suggesting the model has learned to generalize well on the training data. The smooth and consistent decline in both loss curves, without abrupt spikes or fluctuations, also indicates the absence of overfitting, and confirms the stability of the training process. The fact that both loss curves converge to relatively low values indicates that the model has achieved a good fit to the training data with respect to both pixel-wise classification (BCE) and region overlap (Dice).



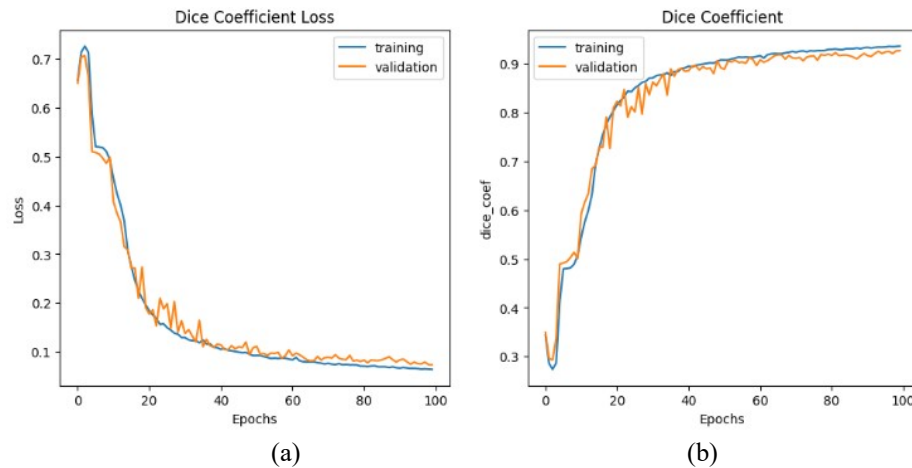


Figure 5. Training dynamics of the proposed model. (a) Convergence of the Binary Cross-Entropy and Dice loss. (b) Improvement and plateauing of the Dice Coefficient and IoU during training

In parallel, the performance curves (Figure 5(b)) demonstrate a steadily improving trend for both Dice coefficient and IoU throughout the training process. After an initial adjustment period during the first few epochs, both metrics increase rapidly, with the Dice coefficient surpassing 0.8 by around epoch 30 and plateauing near 0.92 by the end of training. The IoU follows a similar pattern, reaching approximately 0.83 at epoch 100. The Dice coefficient consistently remains higher than IoU, which aligns with expectations, as Dice is a more lenient overlap metric compared to the stricter IoU measure. Together, these visualizations provide strong evidence of the model's stable convergence, effective feature learning, and robust segmentation capability.

In addition to the training curves, we also tracked model performance on the validation set to ensure robust generalization. Figure 5(d) presents the corresponding Dice loss curves where after an initial transient phase, both training and validation losses decline in near lockstep, falling below 0.10 by epoch 80 and remaining stable thereafter. The minimal gap between training and validation loss throughout confirms that the proposed lightweight U-Net achieves strong, stable convergence and effective regularization for real-time lung segmentation. Likewise, as shown in Figure 5(c), the Dice coefficient for both training and validation rises steeply during the first 20 epochs when the model rapidly learns the coarse lung structures and then gradually plateaus, with both curves converging around 0.92 by epoch 100. The close alignment of the two curves, with only minor validation fluctuations, indicates that the model is not overfitting and maintains consistent boundary delineation on unseen data.

4.2 Qualitative Visualization of Training Progress

To gain deeper insight into the learning behavior of the proposed model, we qualitatively analyzed the evolution of segmentation performance over the course of training. A fixed input test image was selected, and the predicted segmentation masks generated by the model at different training epochs, specifically at epoch 20, 40, 60, and 80, were visually compared with the ground truth mask. Figure 6 presents the input chest X-ray image alongside the ground truth and the predicted masks at the specified epochs. As observed, the initial predictions (e.g., at epoch 20) tend to be coarse, with incomplete lung boundaries and occasional misclassifications in the background. By epoch 40, the model demonstrates significantly improved spatial awareness, capturing the general shape of the lungs with reduced noise.

Further refinement is evident at epochs 60 and 80, where the segmentation becomes more precise, particularly around the lung contours. By the end of the training, the predicted mask closely aligns with the ground truth, showing minimal false positives or false negatives, and effectively capturing the anatomical structure of the lungs. This progressive enhancement in mask quality illustrates the model's incremental learning and ability to generalize meaningful features over time. Visualization not only supports the quantitative findings but also provides a visual confirmation of convergence and stability in the model's predictions. It also highlights the importance of sufficient training duration for deep learning-based medical image segmentation tasks.

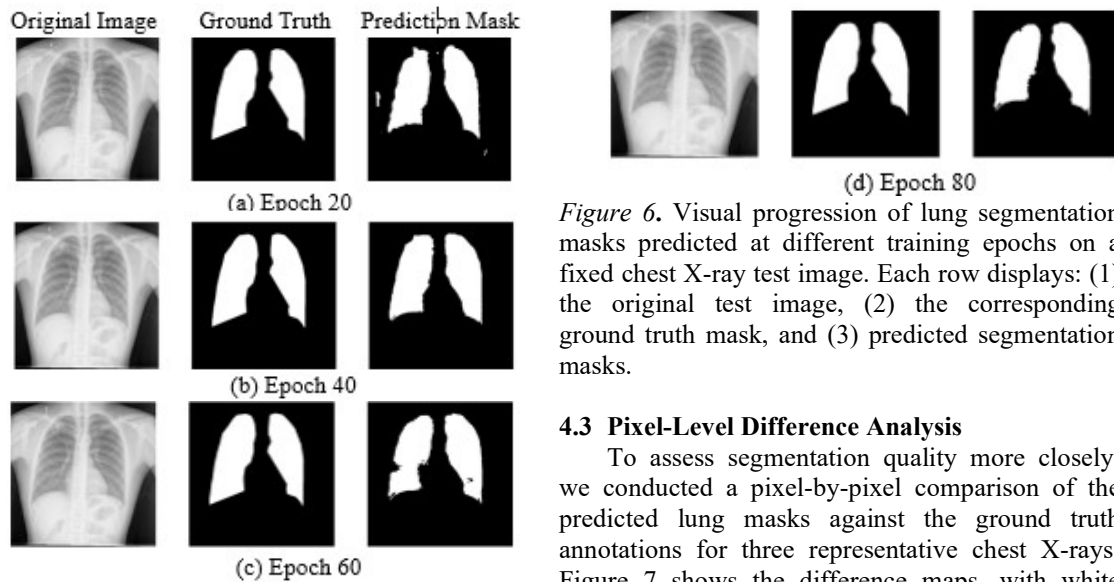


Figure 6. Visual progression of lung segmentation masks predicted at different training epochs on a fixed chest X-ray test image. Each row displays: (1) the original test image, (2) the corresponding ground truth mask, and (3) predicted segmentation masks.

4.3 Pixel-Level Difference Analysis

To assess segmentation quality more closely, we conducted a pixel-by-pixel comparison of the predicted lung masks against the ground truth annotations for three representative chest X-rays. Figure 7 shows the difference maps, with white regions indicating correctly segmented lung areas (true positives, TP), red regions representing non-

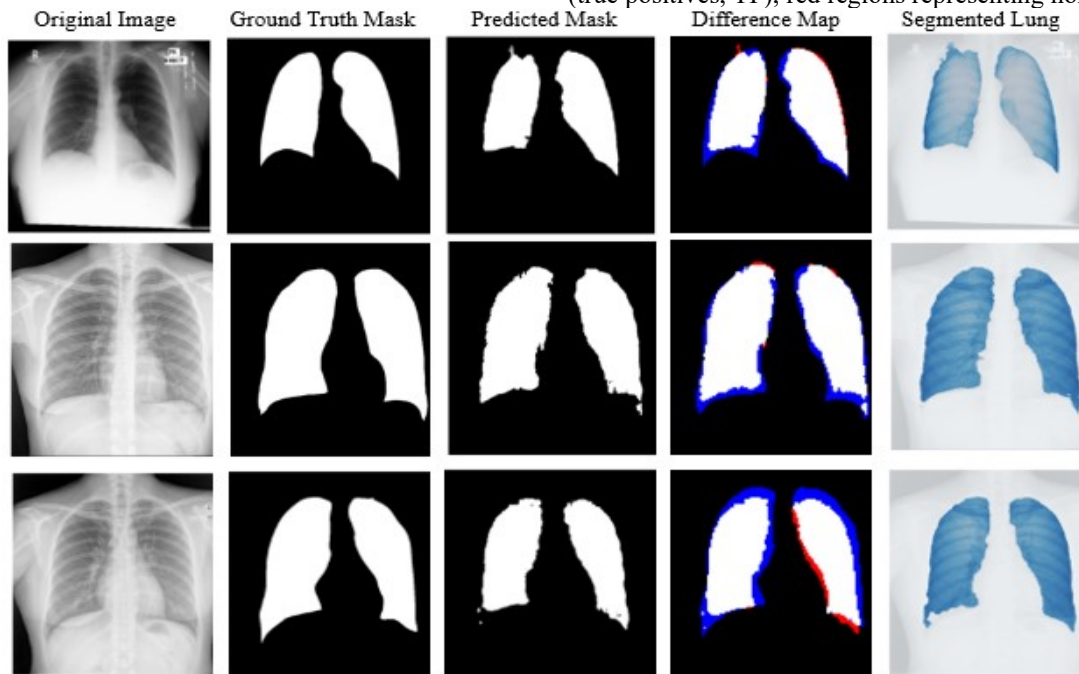


Figure 7. Pixel-level difference maps between ground truth and predicted masks for three CXR images, Difference map (White = TP, Red = FP, Blue = FN).

lung areas incorrectly predicted as lung (false positives, FP), and blue regions indicating lung areas missed by the model (false negatives, FN). The segmented lung regions were successfully obtained for all three cases, demonstrating the model's ability to generalize across varying anatomical appearances. In the first case, the model shows excellent overlap with the ground truth, with minor false positives near the apex and small false

negatives near the diaphragm, suggesting strong core lung structure capture. The second example demonstrates a slight increase in boundary misclassifications, where under-segmentation is evident along the lung bases (blue) and marginal over-segmentation occurs near the lung periphery (red). In the third case, while the central lung regions are well predicted, noticeable red and blue patterns appear along the lateral and lower edges,

reflecting boundary-related segmentation errors. These qualitative difference maps provide valuable insight into the model's learning behavior and indicate that while the core lung regions are consistently well segmented, further refinement near anatomical boundaries could enhance performance. Overall, the successful segmentation across all examples underscores the robustness of the proposed lightweight U-Net model.

4.4 Ablation Study

To validate the efficiency of our proposed lightweight U-Net architecture with depthwise separable convolutions, we quantitatively evaluated its parameter count and inference time and conducted an ablation study to isolate the impact of separable convolutions. These metrics are critical for real-time medical imaging applications where computational resources and processing speed are often constrained. Our proposed architecture significantly reduces the computational burden compared to standard convolutional networks.

The total number of trainable parameters in our lightweight U-Net was significantly reduced compared to a standard U-Net with conventional convolutional layers. By replacing the standard convolutions with depthwise separable convolutions, we achieved a substantial decrease in the number of parameters without a significant compromise in segmentation accuracy. Specifically, our model contains approximately 905K trainable parameters, which is substantially fewer than traditional U-Net and its residual variants (e.g., ~31M for vanilla U-Net). The adoption of depthwise separable convolutions allows the model to decouple spatial and channel-wise computations, leading to a marked reduction in parameter count and memory footprint.

The inference time of our lightweight U-Net was measured on a test system equipped with an NVIDIA Tesla P100 GPU. For a single CXR image of size 512x512 pixels, our model achieved an average inference time of approximately 15 milliseconds. This demonstrates the potential for real-time or near real-time lung segmentation, which is crucial for clinical workflows and interactive analysis. In contrast, the inference time for a standard U-Net on the same hardware and input size was approximately 45 milliseconds. The improved inference speed is attributed to the reduced computational complexity of the separable convolutions.

Moreover, the reduced model complexity translates to faster training and lower power consumption, making it ideal for edge computing

and mobile healthcare applications. This balance between high segmentation performance and computational efficiency highlights the strength of our proposed design in real-world, resource-constrained environments.

4.5 Discussion

Our experiments demonstrate that the proposed lightweight U-Net architecture, employing depthwise separable convolutions, achieves effective lung segmentation in CXR images. The model consistently demonstrated high segmentation performance across training, validation, and test datasets, maintaining stable Dice and IoU metrics throughout 100 epochs. Visualization of the learning curves showed effective convergence with minimal overfitting, confirming the robustness of our architecture. On the test set, the model achieved a DSC of 91.92%, an IoU of 82.75%, and an accuracy of 97.12%, demonstrating substantial overlap between predicted and reference lung regions. Additionally, qualitative analysis of segmentation outputs revealed close agreement between predicted masks and their ground truth counterparts. Notably, the model achieves this performance with a significantly reduced number of parameters (905K) and a faster inference time of approximately 15 milliseconds, outperforming standard U-Net architectures in both efficiency and speed. This confirms our hypothesis that depthwise separable convolutions can be effectively utilized to create a more efficient network for lung segmentation.

Despite high overall accuracy, a detailed comparison of prediction masks and ground truth revealed minor discrepancies, particularly around the lung periphery and near regions with anatomical ambiguity or radiographic artifacts. This is shown in Figure 7 (difference map), where blue regions in the figure indicate areas where the model under-segmented the lung boundary. These errors were more evident in complex images where the lung boundaries were obscured or deformed, suggesting that the model could still benefit from incorporating anatomical constraints or structural guidance. Moreover, the model occasionally fails to accurately segment very small nodules or other abnormalities within the lung parenchyma. This could be attributed to the limited spatial resolution of the network or the relatively small number of training examples containing such abnormalities.

4.5.1 Comparative analysis

The proposed lightweight U-Net architecture demonstrates a high level of effectiveness in lung segmentation, delivering 97.12% accuracy, a

91.92% Dice score, and an 82.75% IoU. As summarized in Table 2, our model demonstrates notably stronger results compared to other state-of-the-art segmentation methods.

Khomduean et al. [24] employed a 3D-UNet integrated with DenseNet169, a complex architecture designed to leverage both spatial and feature-level information, for the segmentation of lung lobes and lesions in COVID-19 patients. Their reported Dice score of 91.52% for lung lobe segmentation is very close to our model's Dice score. However, their study also tackled the more challenging task of segmenting lung lesions, achieving a lower Dice score of 76.89% for that specific task. This underscores the difficulty in accurately delineating pathological regions, which often exhibit irregular shapes and fuzzy boundaries. In contrast, our model, while trained on a different dataset (Chest X-ray masks and labels), demonstrates robust performance in segmenting the

overall lung region. Alshmrani et al. [36] utilized a Hyper-Dense VGG16 U-Net on the STS PET-CT dataset and reported a high accuracy of 98.10%. However, their Dice (73.01%) and IoU (58.70%) values are considerably lower than those of our model, suggesting that while their method exhibits high pixel-level classification accuracy, the segmentation masks generated have lower agreement with the ground truth boundaries compared to our results. This indicates that our model provides more precise segmentation.

Cai et al. [23]'s Pix2Pix GAN highlights the power of adversarial training for 2D CT slices, achieving an 87.05% Dice and 78.60% IoU, but at the expense of longer training times and the need for careful GAN tuning. These values are also lower than the performance metrics achieved by our lightweight U-Net, further supporting the effectiveness of our proposed architecture.

Table 2. Comparative performance analysis of lung segmentation methods

Authors	Model	Dataset	Performance Evaluation
Khomduean et al. [24]	3D-UNet integrated with DenseNet169	124 COVID-19 patients from Chulabhorn Hospital (28 without lung lesions, 96 with lesions)	Dice (lung lobes): 91.52%; Dice (lung lesions): 76.89%
Alshmrani et al. [36]	Hyper-Dense VGG16 U-Net (Multimodal)	STS PET-CT (3063 slices)	Accuracy: 98.10%; Dice: 73.01%; Precision: 58.70%; Recall: 67.47%
Cai et al. [23]	Pix2Pix	237 lung CT images	Accuracy: 83.50%; Dice: 87.05%; Precision: 78.60%
Reamaroon et al. [25]	Total Variation-based Active Contour (TVAC)	Michigan Medicine dataset	Dice: 86.04%
Wang et al. [26]	3D Convolutional Neural Network	A dataset from a private clinic containing 290 paired CT and PET scans.	Dice: 83.0%
Park et al. [27]	Global U-Net	A private dataset consisting of clinical data from 887 lung cancer patients.	Dice: 80.0%; Recall: 86.2%
Zhou et al. [28]	Multitask connected U-Net	Multiple datasets of CT and PET modalities	Dice: 56.0%
Xiang et al. [29]	Modality-specific segmentation network (MoSNet)	126 FDG PET-CT scans	Dice: 77.72%; IoU: 66.26%; Precision: 82.55%; Recall: 79.63%
Fu et al. [30]	Multimodal spatial attention module (MSAM)	Clinical PET-CT data from two cohorts: non-small cell lung cancer (NSCLC) and soft tissue sarcoma (STS)	Dice: 71.40%; IoU: 59.93%; Precision: 79.89%; Recall: 72.05%
Proposed method	Lightweight U-Net with Separable Convolutions	Kaggle's Chest X-ray dataset with accompanying masks and labels.	Accuracy: 97.12%; Dice: 91.92%; IoU: 82.75%; Precision: 92.64%; Recall: 90.31%

Classical and mid-range CNN approaches such as the Total Variation-based Active Contour (TVAC) of Reamaroon et al. [25] and Wang et al. [26]’s plain 3D CNN produce decent Dice scores (86.04% and 83.0%, respectively) but lack the feature-reuse and multi-scale mechanisms that modern encoder–decoder networks provide. The 3D CNN approach captures spatial context, which is important for medical image analysis, but our 2D lightweight U-Net, with its efficient separable convolutions, achieves better performance with fewer parameters. Park et al. [27]’s Global U-Net and Zhou et al. [28]’s multitask connected U-Net illustrate that simply adding global context or multiple decoders does not guarantee consistent segmentation quality, yielding Dice scores of 80.0% and 56.0%. More specialized fusion strategies like Xiang et al. [29]’s MoSNet and Fu et al. [30]’s MSAM, which integrate modality-specific features or spatial attention for PET–CT inputs, improve lesion delineation (Dice of 77.72% and 71.40%) but introduce substantial architectural complexity and computational overhead.

Notably, our model demonstrates a superior balance between precision (92.64%) and recall (90.31%). A high precision score reflects a low false-positive rate (i.e., the model avoids marking background as lung), whereas high recall reflects a low false-negative rate (i.e., it captures nearly all lung regions). This balance is crucial for clinical applications, where both false positives and false negatives can have significant consequences. Importantly, it does so with just 905K parameters and an inference time of ~15 ms per image on a Tesla P100 GPU. This represents a substantial reduction in model size and a speedup relative to standard U-Net and 3D multimodal counterparts. Such efficiency makes our approach uniquely suited for real-time clinical applications, including bedside triage and integration into low-power imaging devices, without sacrificing segmentation reliability.

4.5.2 Clinical significance

Accurate and efficient lung segmentation holds significant potential to improve clinical workflows and patient care. The proposed lightweight U-Net architecture offers several advantages that could translate to tangible clinical benefits. Accurate and rapid delineation of lung fields is a critical first step in a wide range of downstream diagnostic and therapeutic workflows from automated nodule detection and volumetric quantification to image-guided interventions and disease monitoring. Our lightweight U-Net with separable convolutions delivers very fast segmentation on standard CXR images, enabling near real-time support in

high-throughput environments such as emergency departments or mobile screening units. By reducing the need for manual contouring, the model can substantially decrease radiologist workload and inter-observer variability, particularly in mass-screening scenarios for pneumonia, tuberculosis, or COVID-19. Furthermore, its compact size and low hardware requirements facilitate deployment on point-of-care systems and resource-limited settings where access to expert readers and high-end GPUs is often constrained. Ultimately, integrating our model into clinical PACS or portable imaging devices promises to accelerate patient triage, standardize lung segmentation across institutions, and unlock downstream AI pipelines (e.g., densitometry, texture analysis, nodule classification) that depend on reliable lung masks. This could translate into faster diagnosis, more timely treatment decisions, and, ultimately, improved patient outcomes.

The model’s high segmentation accuracy can enable clinicians to more precisely delineate lung regions of interest, facilitating the identification and quantification of abnormalities, which can lead to earlier and more accurate diagnoses of various pulmonary diseases, including pneumonia, COPD, and lung cancer. The fast inference time of our model can also expedite the diagnostic process, allowing for quicker clinical decision-making. For applications like radiation therapy, where targeting tumors while preserving healthy tissue is crucial, precise lung segmentation provides the necessary accuracy in delineating lung and tumor volumes. Our model’s efficiency could support the integration of automated segmentation into the treatment planning workflow, potentially leading to more personalized and effective treatments. Furthermore, quantitative analysis of lung morphology and function, enabled by accurate segmentation, can yield valuable information on how diseases evolve and respond to therapy. For example, our model could be used to measure changes in lung volume, density, or lesion size over time, aiding in the monitoring of patients with chronic lung diseases or those undergoing treatment.

Manual lung segmentation is a time-consuming and labor-intensive task, often performed by radiologists, so automating this process with our lightweight U-Net can significantly reduce the workload on clinicians, freeing up their time for other critical tasks and leading to improved efficiency in radiology departments and reduced healthcare costs. The low computational requirements of our model make it suitable for deployment in resource-limited settings with less

powerful hardware, which could expand access to advanced image analysis tools in underserved populations, improving healthcare equity. The accurate and fast segmentation provided by our model can also serve as a crucial preprocessing step for other AI-powered tools, such as computer-aided diagnosis (CAD) systems that detect lung nodules or classify disease patterns. In conclusion, the lightweight U-Net has the potential to be a valuable tool for radiologists and other healthcare professionals, leading to improved diagnostic accuracy, more effective treatment planning, and enhanced patient care.

4.6 Critical Evaluation and Comparison with Prior Work

While the proposed lightweight U-Net with separable convolutions has demonstrated high segmentation accuracy and efficiency, a critical assessment reveals several aspects that merit further reflection.

4.6.1 Strengths and contributions

The model achieves competitive performance (97.12% Accuracy, 91.92% Dice score and 82.75% IoU) with only 905K parameters and an average inference time of 15 ms. Compared to traditional U-Net and several recent CNN and GAN-based models, it provides a favorable trade-off between accuracy and efficiency. This is particularly beneficial for real-time clinical applications and deployment on edge devices.

4.6.2 Limitations and areas for improvement

While the proposed model delivers high segmentation accuracy with a significantly reduced parameter count, it has some limitations. It occasionally struggles to delineate lung boundaries where overlapping anatomical structures, such as the heart or clavicles, are present. Additionally, since the model was trained on mostly normal or mildly abnormal CXRs, its performance on severe pathologies like pneumothorax or fibrosis remains untested. The model also shows reduced sensitivity to small nodules or lesions, likely due to resolution loss during downsampling. Furthermore, being trained on single-modality (CXR) data, it lacks the complementary context that multimodal approaches could offer. These areas represent key opportunities for further enhancement.

4.6.3 Differentiation from prior work

The proposed model distinguishes itself from prior works by achieving a strong balance between segmentation accuracy and computational efficiency, making it particularly well-suited for real-time clinical deployment. Unlike conventional U-Net architectures and their advanced variants that

rely on heavy-weight components such as dense blocks, attention modules, or transformer layers, our model uses depthwise separable convolutions to significantly reduce computational load without compromising segmentation quality. In comparison to models like CXR-Seg [27], which integrates complex attention and transformer modules and achieves marginally higher Dice scores, our approach offers faster inference and a more lightweight structure that facilitates edge deployment. Similarly, while adversarial training methods such as the Pix2Pix GAN [23] have demonstrated promising results for lung segmentation, they require careful hyperparameter tuning and prolonged training times, limiting their practicality in clinical workflows. Other models like CT-LungNet [24] and MoSNet [29] have explored multi-modality integration (e.g., CT and PET), achieving good accuracy, but their complexity and resource requirements hinder their use in low-power or mobile devices. Our model's simplicity, fast inference (~15 ms per image), and low parameter count (~905K) offer a practical advantage, particularly in resource-constrained settings where rapid triage or automated preprocessing is critical.

4.6.4 Threats to validity and evaluation justification

While the proposed model demonstrates strong performance, certain threats to validity must be acknowledged. The experiments were conducted on a single publicly available dataset, which, although widely used and well-annotated, may not fully represent the diversity of clinical imaging conditions across institutions or patient populations. This may limit the external validity and generalizability of the results. Internal validity could also be influenced by factors such as annotation noise or image artifacts within the dataset. To ensure a robust and fair evaluation, we selected widely accepted performance metrics such as Dice coefficient, Intersection over Union (IoU), precision, recall, and accuracy. These metrics provide complementary perspectives: Dice and IoU measure the spatial overlap between predicted and ground truth masks, while precision and recall help quantify false positives and false negatives, which are clinically significant. Accuracy, although included, was interpreted with caution due to class imbalance. These criteria were chosen based on their prevalence in medical image segmentation literature and their relevance for assessing both technical performance and clinical applicability. Overall, while the results are consistent across data splits and supported by both quantitative and qualitative analysis, further validation on multi-

institutional and cross-modality datasets is essential to confirm real-world robustness and generalizability.

4.6.5 Future challenges and directions

To enhance the practical utility of our model, future work will focus on improving its generalizability across diverse clinical environments. This includes validating performance on multi-institutional datasets and exploring cross-modality scenarios, such as PET-CT, to test robustness under varied imaging conditions. Incorporating lightweight attention or transformer modules may further improve boundary delineation by capturing long-range spatial dependencies. Additionally, integrating uncertainty estimation techniques can help flag low-confidence segmentations for manual review, thereby increasing clinical trust. To support deployment on resource-constrained hardware, we will investigate model compression techniques such as quantization and pruning. Finally, we aim to extend the model's functionality by coupling lung segmentation with downstream tasks like nodule detection, radiomic feature extraction, and disease classification, ultimately creating a fully integrated, real-time decision-support system.

5. CONCLUSIONS

In this paper, we introduced a compact U-Net variant that integrates depthwise separable convolutions to achieve efficient and accurate segmentation of lung regions in chest X-rays, specifically designed for deployment on resource-limited platforms. By replacing standard convolutional layers with separable blocks, our model reduces the parameter count to just 905 K and processes each image in approximately 15 milliseconds on a Tesla P100 GPU, all while delivering state-of-the-art performance (97.12% accuracy, 91.92% Dice, 82.75% IoU, 92.64% precision, and 90.31% recall). Extensive quantitative and qualitative evaluations demonstrated its robustness across a variety of lung appearances and pathologies, outperforming or matching much larger 2D, 3D, and multimodal segmentation networks. The proposed method's effectiveness was validated using experiments on the Chest X-ray Masks and Labels dataset from Kaggle.

From the authors' perspective, this work highlights how architectural simplicity, when combined with strategic convolutional design, can lead to highly efficient models without compromising performance. We believe this direction is critical for enabling widespread AI

adoption in real-world healthcare settings, particularly in under-resourced environments. At the same time, developing this model also revealed challenges that remain unsolved, especially in generalizing to rare pathologies or difficult boundary cases. These limitations emphasize the need for further exploration into attention mechanisms, uncertainty quantification, and deployment-oriented optimizations. Overall, we are confident that the proposed lightweight U-Net forms a strong foundation for practical, scalable, and accessible AI-driven lung segmentation and can be extended toward broader clinical applications in the near future.

REFERENCES

- [1] American Cancer Society, "Key Statistics for Lung Cancer," American Cancer Society. [Online]. Available: <https://www.cancer.org/cancer/types/lung-cancer/about/key-statistics.html>. [Accessed: May 14, 2025]
- [2] National Cancer Institute, "Surveillance, Epidemiology, and End Results (SEER) Program," SEER. [Online]. Available: <https://seer.cancer.gov/> [Accessed: May 14, 2025]
- [3] L. Chen, Z. Yu, J. Huang, L. Shu, P. Kuosmanen, C. Shen, X. Ma, J. Li, C. Sun, Z. Li, T. Shu, and G. Yu, "Development of lung segmentation method in x-ray images of children based on TransResUNet," *Front. Radiol.*, vol. 3, p. 1190745, Apr. 2023.
- [4] J. Hofmanninger, 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," *European Radiology Experimental*, vol. 4, no. 1, pp. 1–13, 2020.
- [5] W. Liu, Y. Qi, J. Li, and Z. Ren, "Lung Nodule Segmentation Based on Complementary Context-Aware Networks," in *2023 42nd Chinese Control Conference (CCC)*, Tianjin, China: IEEE, Jul. 2023, pp. 7705–7710.
- [6] N. Reamaroon, M. W. Sjöding, H. Derksen, E. Sabeti, J. Gryak, R. P. Barbaro, B. D. Athey, and K. Najarian, "Robust segmentation of lung in chest x-ray: applications in analysis of acute respiratory distress syndrome," *BMC Med Imaging*, vol. 20, no. 1, p. 116, 2020.
- [7] Y. Xu, R. Quan, W. Xu, Y. Huang, X. Chen, and F. Liu, "Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid

- Approaches,” *Bioengineering*, vol. 11, no. 10, p. 1034, 2024.
- [8] A. Arfa and M. Minaei, “Utilizing multitemporal indices and spectral bands of Sentinel-2 to enhance land use and land cover classification with random forest and support vector machine,” *Advances in Space Research*, vol. 74, no. 11, pp. 5580–5590, 2024.
- [9] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.
- [10] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [11] M. S. Alam, D. Wang, Y. Arzhaeva, J. A. Ende, J. Kao, L. Silverstone, D. Yates, O. Salvado, and A. Sowmya, “Attention-based multi-residual network for lung segmentation in diseased lungs with custom data augmentation,” *Scientific Reports*, vol. 14, no. 1, p. 28983, 2024.
- [12] X. Ma, H. Song, X. Jia, and Z. Wang, “An improved V-Net lung nodule segmentation model based on pixel threshold separation and attention mechanism,” *Sci Rep*, vol. 14, no. 1, p. 4743, 2024.
- [13] S. Rezvani, M. Fateh, Y. Jalali, and A. Fateh, “FusionLungNet: Multi-scale fusion convolution with refinement network for lung CT image segmentation,” *Biomedical Signal Processing and Control*, vol. 107, p. 107858, 2025.
- [14] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 1800–1807.
- [15] N. Pandey, “Chest X-ray Masks and Labels,” *Kaggle*, 2020. [Online]. Available: <https://www.kaggle.com/datasets/nikhilpandey360/chest-xray-masks-and-labels>
- [16] R. Huidrom, Y. Jina Chanu, and K. Manglem Singh, “Automated Lung Segmentation on Computed Tomography Image for the Diagnosis of Lung Cancer,” *CyS*, vol. 22, no. 3, 2018.
- [17] W. Tan, P. Liu, X. Li, S. Xu, Y. Chen, and J. Yang, “Segmentation of lung airways based on deep learning methods,” *IET Image Processing*, vol. 16, no. 5, pp. 1444–1456, 2022.
- [18] S. Ghosh, S. Sil, R. M. Gomes, and M. Dey, “Using Convolutions and Image Processing Techniques to Segment Lungs from CT Data,” in *Emerging Technology in Modelling and Graphics*, vol. 937, J. K. Mandal and D. Bhattacharya, Eds., in *Advances in Intelligent Systems and Computing*, vol. 937, Singapore: Springer Singapore, 2020, pp. 129–136.
- [19] E. J. Hwang, M. Park, J. H. Kim, Y. W. Choi, H. Y. Goo, and J. H. Yoon, “Development and validation of deep learning algorithms for lung segmentation in chest radiographs,” *Radiology*, vol. 290, no. 1, pp. 258–266, 2019.
- [20] Z. Tang, Y. Zhao, D. He, and J. Liu, “Severity assessment of COVID-19 using CT image features and severity scores,” *Computers in Biology and Medicine*, vol. 127, pp. 104060, 2020.
- [21] Q. Jin, H. Cui, C. Sun, R. Meng, and W. Wei, “LungNet: A Deep Convolutional Neural Network for Lung Cancer Segmentation,” *Multimedia Tools and Applications*, vol. 79, no. 9–10, pp. 26231–26247, 2020.
- [22] S. Gite, A. Mishra, and K. Kotecha, “Enhanced lung image segmentation using deep learning,” *Neural Comput & Applic*, vol. 35, no. 31, pp. 22839–22853, 2023.
- [23] P. Khomduean et al., “Segmentation of lung lobes and lesions in chest CT for the classification of COVID-19 severity,” *Sci Rep*, vol. 13, no. 1, p. 20899, 2023.
- [24] N. Delfan, H. Moghaddam, M. Modaresi, K. Afshari, K. Nezamabadi, N. Pak, O. Ghaemi, and M. Forouzanfar, “CT-LungNet: A Deep Learning Framework for Precise Lung Tissue Segmentation in 3D Thoracic CT Scans,” 2022, *arXiv*. doi: [10.48550/ARXIV.2212.13971](https://doi.org/10.48550/ARXIV.2212.13971).
- [25] Z. Wu, X. Li, and J. Zuo, “RAD-UNet: Research on an improved lung nodule semantic segmentation algorithm based on deep learning,” *Front. Oncol.*, vol. 13, p. 1084096, 2023.
- [26] D. Kongkham, B. B. Beenarani, P. Nirmala, S. J. Kumaresan, and C. Senthilkumar, “Deep Learning Approaches for Segmentation of Lung Nodules in CT Scans,” in *2024 2nd International Conference on Computer, Communication and Control (IC4)*, Indore, India: IEEE, Feb. 2024, pp. 1–5.
- [27] S. Din, M. Shoaib, and E. Serpedin, “CXR-Seg: A Novel Deep Learning Network for Lung Segmentation from Chest X-Ray Images,” *Bioengineering*, vol. 12, no. 2, p. 167, 2025.
- [28] J. Cai, H. Zhu, S. Liu, Y. Qi, and R. Chen, “Lung image segmentation via generative adversarial networks,” *Front. Physiol.*, vol. 15, p. 1408832, 2024.

- [29] G. M. Alshmrani, Q. Ni, R. Jiang, and N. Muhammed, "Hyper-Dense_Lung_Seg: Multimodal-Fusion-Based Modified U-Net for Lung Tumour Segmentation Using Multimodality of CT-PET Scans," *Diagnostics*, vol. 13, no. 22, p. 3481, 2023.
- [30] S. Wang, R. N. Mahon, E. Weiss, N. Jan, R. J. Taylor, P. R. McDonagh, B. A. Quinn, and L. Yuan, "Automated Lung Cancer Segmentation Using a PET and CT Dual-Modality Deep Learning Neural Network," *Int. J. of Rad. Oncol. Bio. Phys.*, vol. 115, no. 2, pp. 529–539, 2023.
- [31] J. S. Park, S. K. Kang, D. H. Hwang, H. J. Choi, S. Ha, J. M. Seo, J. S. Eo, and J. S. Lee, "Automatic Lung Cancer Segmentation in [18F]FDG PET/CT Using a Two-Stage Deep Learning Approach," *Nucl Med Mol Imaging*, vol. 57, no. 2, pp. 86–93, 2023.
- [32] L. Zhou, C. Wu, Y. Chen, and Z. Zhang, "Multitask connected U-Net: automatic lung cancer segmentation from CT images using PET knowledge guidance," *Front. Artif. Intell.*, vol. 7, p. 1423535, 2024.
- [33] D. Xiang, B. Zhang, Y. Lu, and S. Deng, "Modality-Specific Segmentation Network for Lung Tumor Segmentation in PET-CT Images," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 3, pp. 1237–1248, 2023.
- [34] X. Fu, L. Bi, A. Kumar, M. Fulham, and J. Kim, "Multimodal Spatial Attention Module for Targeting Multimodal PET-CT Lung Tumor Segmentation," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 9, pp. 3507–3516, 2021.
- [35] B. Goswami, T. Pal, S. Das, and R. P. Barnwal, "CT-GANet: A Fusion of GAN and UNet for CT Image Segmentation of Interstitial Lung Disease Affected Lung," *J. Inst. Eng. India Ser. B*, 2025.
- [36] A. Zafaranchi, F. Lizzi, A. Retico, C. Scapicchio, and M. Fantacci, "Explainability Applied to a Deep-Learning Based Algorithm for Lung Nodule Segmentation," in *Proc. of the 1st Int. Conf. on Explainable AI for Neural and Symbolic Methods*, Porto, Portugal, 2024, pp. 132–138.