# Pneumothorax Segmentation In Chest X-Rays Using UNet++ And EfficientNet

Xavier Tan Ying Le
*Faculty of Engineering, Nanyang Polytechnic, Singapore*
Email: xaviertanyingle1@gmail.com

**Abstract—Pneumothorax, commonly known as collapsed lung, occurs when air enters the pleural space between the chest wall and lung. While chest X-rays are the primary diagnostic tool, accurate interpretation requires experienced radiologists, who may not be readily available in under-resourced areas. This paper presents an enhanced deep learning approach for automated pneumothorax segmentation in chest X-rays. I propose a modified UEfficientNet architecture with layer-specific dropout and advanced preprocessing techniques, including Contrast Limited Adaptive Histogram Equalization (CLAHE) and intelligent lung area cropping. Using the SIIM-ACR Pneumothorax Segmentation Challenge dataset, the model incorporates transfer learning with EfficientNetB4 pre-trained on ImageNet, enhanced data augmentation, and a weighted loss function combining binary cross-entropy and Dice loss to address class imbalance.**

**Index Terms—Pneumothorax segmentation, BioMedical Image Processing, Deep Learning, UEfficientNet, Data Augmentation, Test-Time Augmentation, Chest X-ray, Transfer Learning**

## I. Introduction

Pneumothorax represents a critical medical condition where air accumulates in the pleural space, compromising lung function and potentially leading to respiratory distress [1]. The condition demands prompt diagnosis and intervention, as delayed treatment can result in severe complications. Currently, chest X-rays serve as the primary diagnostic tool due to their accessibility and cost-effectiveness [2]. The diagnosis of pneumothorax through chest radiographs faces several challenges in clinical practice: visual ambiguity in radiographic changes, critical time sensitivity for treatment, and limited access to experienced radiologists in underserved areas. Studies have also shown significant inter-observer variability in pneumothorax detection, even among experienced radiologists [11].

Recent advancements in deep learning have demonstrated remarkable potential in medical image analysis [4]. Convolutional Neural Networks have shown particular promise in automated feature extraction from radiological patterns, while providing standardized interpretation that reduces inter-observer variability. These systems offer rapid processing capabilities and continuous availability, operating without fatigue throughout the day.

This study introduces several key innovations in pneumothorax detection through an enhanced architecture design. I implement a modified UEfficientNet with layer-specific dropout rates optimized for chest radiograph analysis, coupled with advanced preprocessing techniques combining CLAHE enhancement and fixed-region cropping. The training process employs a custom weighted loss function to address class imbalance issues. The approach utilizes the EfficientNetB4 [7] architecture pre-trained on ImageNet as a backbone, enhanced with custom modifications for medical image segmentation. The implementation processes images at 256 x 256 resolution, balancing computational efficiency with diagnostic accuracy.

## II. Related Work

With the growth of deep learning techniques in medical image processing [4], researchers have focused on various challenges of medical image segmentation. Early approaches included the application of CNNs [10] using small kernels for brain tumor segmentation in Magnetic Resonance Imaging (MRI). Remarkable results were achieved in semantic segmentation tasks [4], where each pixel in the image was labeled and classified into its respective category.

### *A. U-Net Development*

A significant breakthrough came with the development of U-Net by Ronneberger et al. [11], an end-to-end semantic segmentation network specifically designed for biomedical images (Fig. 1). The U-Net architecture achieved exceptional performance across numerous biomedical segmentation applications, successfully handling neuronal structure segmentation, cell segmentation in microscopy images, retinal layer segmentation from OCT scans, and brain tumor segmentation from MRI images. The architecture's success stems from its symmetric encoder-decoder structure and the introduction of skip connections, which preserve crucial spatial information during the upsampling process.

### *B. UNet++ Architecture*

Building upon U-Net's success, Zhou et al. introduced UNet++ [5], a more powerful architecture based on nested and dense skip connections (Fig. 2). The architecture introduces several significant improvements through its nested dense skip pathways, which bridge the semantic gap between encoder and decoder while ensuring maximum information flow through dense connections. The nested structure enables deep supervision and supports multi-scale feature fusion, allowing progressive feature integration at different semantic levels. This design enhances feature propagation through multiple paths and improves gradient flow during training. Furthermore, the architecture incorporates deep supervision mechanisms that generate multiple segmentation maps at different scales, enable flexible model pruning at inference time, and improve overall training stability and convergence.

### *C. EfficientNet Development*

The introduction of EfficientNet by Tan and Le [7] marked another milestone in deep learning architecture design. Their work demonstrated that carefully balancing network depth, width, and resolution could lead to both better performance and improved efficiency. The architecture family, particularly EfficientNetB4, has shown remarkable success as a backbone network for various computer vision tasks, including medical image analysis. When compared with traditional architectures like ResNet-50, EfficientNetB4 achieves superior accuracy while maintaining comparable computational efficiency.

### *D. Recent Advances*

In 2019, the SIIM-ACR Pneumothorax Segmentation Challenge [8] provided a platform for developing and comparing automated pneumothorax detection methods. The challenge featured a two-round competition with 1,475 teams participating. In 2020, researchers released a 2ST-UNet [12] architecture for pneumothorax segmentation, utilizing U-Net with ResNet-34 as backbone. Their approach of training from lower to higher resolutions achieved notable results, ranking among the top 9% of solutions.

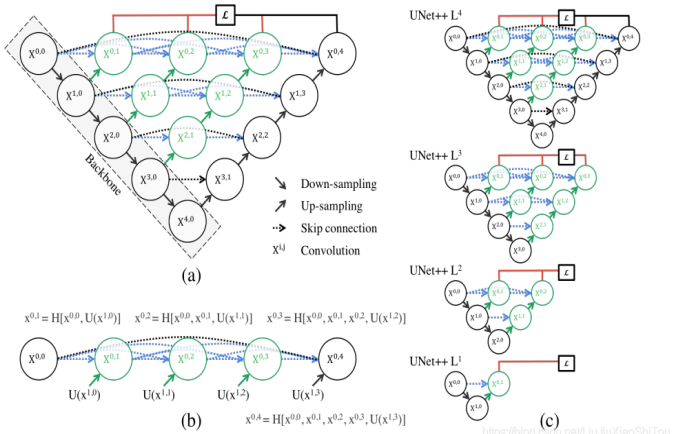


Fig. 1. U-Net architecture [11] showcasing the contracting path (left side) and expansive path (right side) connected by skip connections.

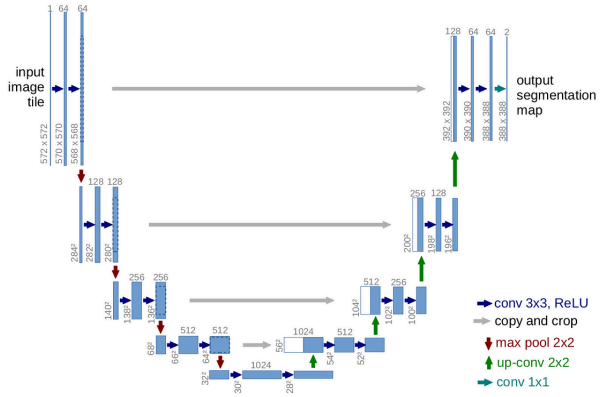


Fig. 2. (a) UNet++ architecture showing the encoder and decoder connected through nested dense convolutional blocks. (b) Detailed analysis of the first skip pathway. (c) Pruning capability at inference time when trained with deep supervision [5].

## III. EXPERIMENTAL INVESTIGATION

### *A. Dataset*

The study utilizes the dataset from the SIIM-ACR Pneumothorax Segmentation Challenge [8], which contains training images and their corresponding masks. All images are provided in Digital Imaging and Communications in Medicine (DICOM) format at 1024 x 1024 pixels resolution. The masks are encoded using Run Length Encoding (RLE) format. For images containing pneumothorax, the RLE serves as a mask value to identify affected pixel locations. Images without pneumothorax are marked with a mask value of -1. The complete dataset includes 12,047 training images with corresponding masks (dataset summary in Table I).

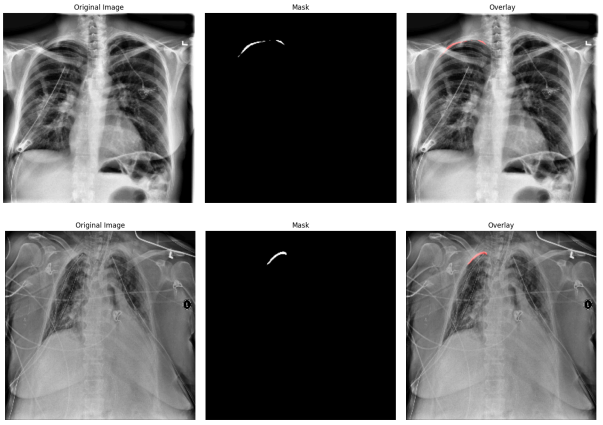


Fig. 3. Sample chest X-ray images containing pneumothorax. Left: Original X-ray image. Middle: Ground truth mask. Right: X-ray with mask overlay.

TABLE I
DATASET SUMMARY

| Training Images | Testing Images | Positive Samples | Masks |
|---|---|---|---|
| 12,047 | 3,205 | 2,669 | 12,954 |

### *B. Experimental Setting*

The implementation was developed using Python 3 with TensorFlow and Keras frameworks. For the primary experiments, images were processed at 256 x 256 pixel resolution using Kaggle's computational resources. The enhanced model architecture incorporated several key components. The model used EfficientNetB4 backbone with pre-trained ImageNet weights, implemented layer-specific dropout rates for optimal regularization, and employed a custom weighted loss function combining binary cross-entropy and Dice loss. The training process utilized the Adam optimizer with an initial learning rate of

0.0001 and Cosine Annealing scheduling. Early stopping with a patience of 15 epochs was implemented to prevent overfitting.

## IV. PROPOSED APPROACH

### *A. Data Pre-processing and Data Augmentation*

The preprocessing pipeline begins with a fixed-coordinate cropping strategy targeting the typical lung area. The cropping boundaries are set at 8% and 78% of image height for top and bottom margins respectively, while horizontal boundaries are established at 10% and 85% of image width, with an additional padding of 20 pixels around the region of interest. The pipeline continues with CLAHE application for contrast enhancement, followed by MinMax scaling for intensity normalization. The data augmentation strategy implements horizontal flipping with 0.3 probability, alongside carefully controlled brightness and contrast adjustments to maintain anatomical validity.

### *B. Model Architecture*

The architecture employs layer-specific dropout rates optimized for different network depths. The bottleneck layers utilize a 0.4 dropout rate for maximum regularization, while middle layers employ rates between 0.2-0.3, and shallow layers maintain a minimal 0.1 rate to preserve essential low-level features. This graduated approach helps balance model capacity with regularization needs across different abstraction levels.

### *C. Training Methodology*

I implement a five-fold cross-validation strategy with dynamic learning rate adjustment through cosine annealing. The batch sizes are optimized for memory efficiency while maintaining training stability. Comprehensive metric monitoring includes both IoU and Dice coefficient tracking throughout the training process. Test-time augmentation (TTA) is employed during validation and inference. The TTA pipeline includes horizontal flipping of input images, generating predictions for both original and augmented versions, then averaging the results after reverse-transforming the augmented predictions. This approach reduces prediction variance and improves model reliability, particularly in cases where the pneumothorax presentation is subtle or ambiguous. The TTA process is integrated into both validation during training and final inference, ensuring consistent evaluation metrics throughout development.

### *D. Loss Function and Implementation Details*

The model employs a weighted combination of binary cross-entropy and Dice loss to address class imbalance in the dataset. The binary cross-entropy component is weighted by class frequencies, with a positive class weight of 3.51 and negative class weight of 1.0, derived from the dataset's class distribution. The Dice loss component helps optimize the segmentation boundary accuracy. The combined loss function is defined as:

$$L = W_{bce} * BCE(y_{true}, y_{pred}) + w_{dice} * (1 - DiceCoeff(y_{true}, y_{pred}))$$

where w_bce and w_dice are the respective weights for each loss component. The training process utilizes the Adam optimizer with an initial learning rate of 0.0001, and implements early stopping with a patience of 15 epochs to prevent overfitting. During inference, test-time augmentation with horizontal flipping is employed to enhance prediction robustness. Model performance is evaluated using the mean Dice coefficient, matching the competition's evaluation metric.

## V. Results and Discussion

### *A. Performance Evaluation*

The model's performance was evaluated using two primary metrics: the Intersection over Union (IoU) and the Dice similarity coefficient (DSC). The IoU metric measures the overlap between predicted and ground truth segmentations, while the DSC provides a measure of overlap that is particularly sensitive to segmentation accuracy. The Dice coefficient is calculated as:

$$\mathbf{2 * |X \cap Y| \; |X| + |Y|}$$

where X represents the predicted set of pixels and Y represents the ground truth. The coefficient is defined as 1 when both X and Y are empty.

### *B. Experimental Results*

The model achieved strong performance across multiple evaluation metrics. In cross-validation, an average training IoU of 0.8996 and validation IoU of 0.8566 was obtained, demonstrating robust generalization capability. The training converged effectively, with final average losses of 0.3000 and 1.1475 for training and validation respectively.

The implementation of test-time augmentation (TTA) yielded significant improvements in prediction accuracy. On a random batch evaluation, TTA increased the correct prediction rate from 62.50% to 75.00%, representing a substantial 12.50 percentage point improvement in accuracy. This enhancement particularly benefited cases with subtle or ambiguous pneumothorax presentations.

Compared to recent approaches in pneumothorax segmentation, the model demonstrates competitive performance. The 2ST-UNet approach [12] achieved notable results in the SIIM-ACR challenge, and the validation IoU of 0.8566 and enhanced accuracy through TTA positions it favorably among current state-of-the-art solutions. The implementation of layer-specific dropout and advanced preprocessing techniques contributes to robust performance across varied clinical presentations.

### *C. Visual Analysis*

Qualitative analysis through visualization (Fig. 4) demonstrates the model's capability to accurately segment pneumothorax regions across various presentations. The results show precise boundary delineation

and robust performance across different pneumothorax sizes and locations. The comparison between ground truth masks and model predictions reveals high fidelity in segmentation, particularly in challenging cases with multiple affected areas.

The training progression (Fig. 5) shows stable convergence across folds, with the IoU metric demonstrating consistent improvement throughout the training process. The loss curves indicate effective learning without significant overfitting, validated by the close tracking between training and validation metrics.

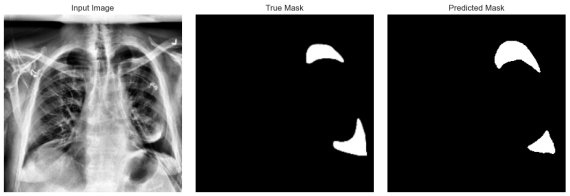


Fig. 4. Visual segmentation results demonstrating model performance. Left: Original chest X-ray input image. Middle: Ground truth mask showing actual pneumothorax regions. Right: Model's predicted segmentation mask.

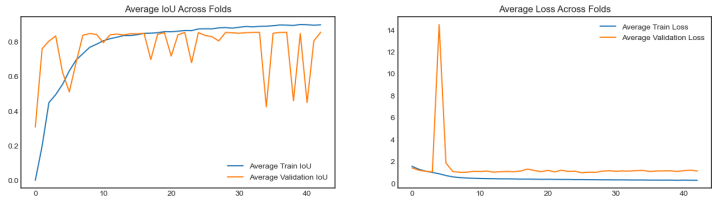


Fig. 5. Training metrics across model development. Left: Average IoU performance across folds showing training (blue) and validation (orange) metrics, demonstrating consistent model convergence. Right: Loss curves for training and validation

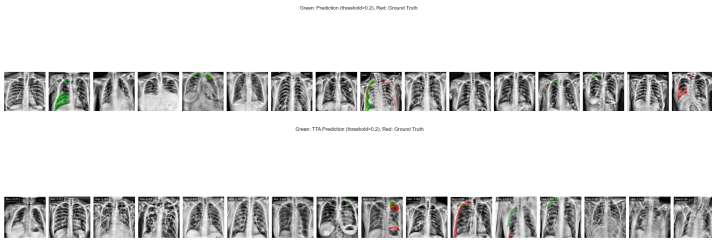

Fig. 6. Comparative analysis of standard predictions versus TTA predictions across diverse cases. Top row: Standard prediction results (green) overlaid with ground truth (red) demonstrating varied pneumothorax presentations and segmentation performance. Bottom row: Test-time augmentation prediction results showing improved segmentation accuracy and reduced false positives compared to standard predictions.

## VI. Conclusion

I presented an enhanced deep learning approach for automated pneumothorax segmentation in chest X-rays. The proposed model employs a modified UEfficientNet architecture with EfficientNetB4 as backbone, pre-trained on the ImageNet dataset. The implementation incorporates several key innovations, including layer-specific dropout rates, advanced preprocessing techniques with CLAHE enhancement, and test-time augmentation.

The model achieved strong performance metrics, with an average validation IoU of 0.8566 and significant improvements in prediction accuracy through test-time augmentation, increasing accuracy from 62.50% to 75.00% on validation sets. The implementation of the comprehensive preprocessing pipeline and weighted loss function effectively addressed the challenges of class imbalance and varying image quality in clinical settings.

While this approach demonstrates strong performance, certain limitations should be noted. The current implementation processes images at 256 x 256 resolution, which may impact the detection of very small pneumothoraces. Additionally, the model's performance might vary when applied to X-rays from different medical institutions or equipment, suggesting the need for broader validation studies.

Future research directions include the exploration of multi-modal imaging integration, development of more sophisticated augmentation strategies, and investigation of real-time processing capabilities for clinical deployment. The promising results achieved by this approach suggest potential for clinical application, particularly in resource-limited settings where expert radiological interpretation may not be readily available.

## References

[1] A. MacDuff, A. Arnold, and J. Harvey, "Management of spontaneous pneumothorax: British Thoracic Society pleural disease guideline 2010," Thorax, vol. 65, no. Suppl 2, pp. ii18-ii31, 2010.
[2] R. Thomas, Y. C. G. Lee, "Management of pleural disease," Clinical Chest Medicine, vol. 34, no. 1, pp. 47-55, 2013.
[3] N. L. Chowdhury et al., "Deep learning for automated pneumothorax detection: A systematic review," Journal of Medical Imaging and Radiation Oncology, vol. 65, no. 4, pp. 432-441, 2021.
[4] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88, 2017.
[5] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," IEEE Transactions on Medical Imaging, vol. 39, no. 6, pp. 1935-1946, 2020.
[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431-3440, 2015.
[7] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. International Conference on Machine Learning (ICML), pp. 6105-6114, 2019.
[8] R. K. Kohli et al., "SIIM-ACR Pneumothorax Segmentation Challenge," Journal of Digital Imaging, vol. 33, no. 2, pp. 506-512, 2020.
[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.
[10] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1240-1251, 2016.
[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks

for biomedical image segmentation," in Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234-241, 2015.
[12] J. Park et al., "2ST-UNet: 2-Stage Training U-Net for Pneumothorax Segmentation in Chest X-rays," IEEE Access, vol. 8, pp. 122055-122063, 2020.