

Deep Learning For Lung Segmentation And Disease Classification In Chest X-Ray Imaging

Anwar Muta Mohammad*, Abirami J., Sangeetha Banerjee

Department of Computer Science and Information System, Birla Institute of Technology and Science, Pilani, India

ABSTRACT

Pulmonary illness is one of the largest contributors to death and disability across the globe, placing it among the top causes of mortality. Because of this, identifying and categorizing lung abnormalities from chest radiographs quickly and correctly is essential for guiding diagnosis and treatment. This paper surveys the current landscape of techniques used to segment and classify lungs in X-ray images, giving particular attention to deep-learning-driven approaches such as Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and architectures that combine both. The discussion covers influential public datasets, methods for preparing and augmenting training data, segmentation strategies (spanning U-Net and its variants, contour-driven techniques, and the alpha-shape algorithm), and frameworks built for distinguishing among multiple conditions — pneumonia, tuberculosis, chronic obstructive pulmonary disease (COPD), pulmonary fibrosis, and COVID-19. The review also touches on federated learning as a way to protect patient privacy in collaborative AI training, multimodal vision-language systems for interpreting chest radiographs, and the ongoing problems of biased datasets and limited model generalizability. Drawing on more than 40 recent publications, the paper reports leading performance benchmarks and points to areas still requiring research before such systems can be reliably used in clinical practice.

Keywords: Chest X-ray, Lung Segmentation, Lung Disease Classification, Deep Learning, Convolutional Neural Network, Vision Transformer, U-Net, Medical Image Analysis, Federated Learning, Multimodal Models.

INTRODUCTION

Diseases affecting the respiratory system are among the most widespread and dangerous health conditions worldwide, with lung disease ranking third among global causes of death — a fact that highlights how badly reliable, automated diagnostic support is needed [1]. Chest X-ray (CXR) imaging remains one of the most accessible and affordable ways to evaluate the lungs initially, but reading these images accurately still depends heavily on the skill of radiologists, professionals who are often in short supply in under-resourced regions.

Deep learning has reshaped how medical images are analyzed, allowing automated systems to match or even outperform human specialists on certain narrowly defined tasks [2], [3]. Two core problems drive research in this space: segmenting the lungs (tracing their boundaries apart from surrounding anatomy in a CXR) and classifying lung disease

(assigning a diagnostic category once the lung region has been isolated). Both tasks come with their own obstacles — imbalanced classes, imaging noise and artifacts, disagreement between different observers, and the fact that pathological findings can look very different from one patient to another [4].

Before deep learning took hold, lung segmentation depended on manually engineered features and rule-based techniques such as active contours, graph-cut algorithms, and atlas registration [5]. These older methods were easy to interpret but fell short when anatomy varied or disease altered the expected shape of the lungs. The U-Net architecture and other fully convolutional networks eventually replaced these classical techniques by learning layered feature representations straight from the training data [6]. Attention-driven models and transformers have since pushed performance further by capturing relationships between distant regions of an image more effectively [7].

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

On the classification side, CNNs trained on sizable radiograph collections produced major performance gains for detecting pneumonia, screening for tuberculosis, and diagnosing COVID-19 [8], [9], [42]. A common strategy to work around the scarcity of labeled medical images is transfer learning — reusing networks such as VGG-16, VGG-19, ResNet, DenseNet, and InceptionNet that were originally trained on ImageNet [10]. Alongside image-based methods, classifying respiratory illness from sound recordings converted into spectrograms and analyzed with Vision Transformers has also become a useful complementary diagnostic approach [11].

The remainder of the paper proceeds as follows: Section II covers background material and related research; Section III reviews segmentation techniques; Section IV covers classification approaches; Section V examines datasets and evaluation methods; Section VI looks at cross-domain and privacy-focused methods; Section VII summarizes findings and remaining challenges; and Section VIII concludes the paper.

2. Background and Related Work

2.1 Classical Image Processing Techniques

Prior to deep learning, segmenting lungs and spotting disease in chest radiographs relied mainly on traditional image-processing and machine-learning tools. Techniques like thresholding, morphological filtering, and edge detection were used to isolate the lung fields, while texture-based descriptors — the Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) — helped characterize tissue [12]. Region-growing and watershed algorithms provided partial automation for segmentation but performed poorly when disease obscured the natural lung boundary [43].

In related detection work involving remote sensing and medical imaging, region-based CNNs (R-CNNs) were paired with handcrafted descriptors such as GLCM and Dominant Rotated Local Binary Patterns (DRLBP) to boost detection accuracy in difficult modalities like Synthetic Aperture Radar (SAR) imagery [13]. This shows that carefully engineered, domain-specific features can meaningfully strengthen neural network performance, especially when annotated data is limited.

2.2 The Emergence of Deep CNNs

CNNs reshaped the field of computer vision, and by extension, medical imaging. Foundational research on large-scale ImageNet classification showed that deep, hierarchical feature learning could surpass human-level accuracy on major recognition benchmarks [14]. Transfer learning subsequently made it possible to adapt these pre-trained networks to medical domains where labeled examples are far scarcer [15].

Within respiratory imaging, CNNs were used to sort chest X-rays into normal versus abnormal categories. Networks including VGG-16, VGG-19, ResNet-50, ResNet-101, and DenseNet-121 became common baseline models for comparison [11], [16]. Augmentation methods — random rotations, flips, crops, brightness changes, and elastic warping — proved important for reducing overfitting and helping models generalize to new patients [17].

CNNs have also found use well outside medical imaging, in areas such as traffic-flow monitoring, self-driving systems, and satellite imagery [18]. For instance, an AI-based traffic-management pipeline using CNNs reported 94.89% accuracy, 97.40% precision, and an F1 score of 96.76% [19] — evidence that architectural advances in one domain often transfer readily to others.

2.3 The Emergence of Deep CNNs

The Transformer architecture, originally built for language processing [20], was later adapted for image understanding. Vision Transformers (ViTs) work by splitting an image into fixed-size patches, converting each into an embedding vector, and applying self-attention layers so the model can capture relationships across the whole image. This design has performed well on major image classification tasks, particularly with large-scale pretraining [21].

Applied to respiratory illness, a ViT-based model using attention mechanisms classified lung sounds represented as spectrograms, reaching 91.04% accuracy across five categories — Normal, Asthma, Pneumonia, COPD, and Lung Fibrosis — when paired with EBU normalization and data augmentation, outperforming CNN baselines like ResNet-101 and VGG-19 [11]. The attention maps

generated by the ViT also offer some interpretability by showing which parts of the spectrogram most influenced the prediction.

3. Lung Segmentation Methods

3.1 U-Net and Fully Convolutional Networks

U-Net [22], developed originally for biomedical segmentation tasks, uses an encoder-decoder design with skip connections that help preserve spatial detail lost during downsampling. It became the standard choice for medical segmentation largely because it performs well even with a modest number of labeled images. Several variants have followed: Attention U-Net, which introduces channel and spatial attention into the skip connections; Residual U-Net, which adds residual blocks to improve gradient flow; and U-Net++, which redesigns the skip pathways using dense feature aggregation [23].

For segmenting lungs in chest X-rays, fully convolutional networks generate pixel-level masks labeling each pixel as either lung tissue or background. Combining features across multiple scales is important since lung boundaries vary widely in size and shape. Typically, a pre-trained backbone such as ResNet-50 or VGG-16 handles feature extraction (taking advantage of ImageNet-based representations), while the decoder restores resolution using transposed convolutions or bilinear upsampling [24].

3.2 Boundary-Driven and Alpha-Shape Segmentation

One newer strategy pairs a semantic boundary-detection network with the alpha-shape algorithm, a geometric technique that constructs a polygon around a set of points based on a chosen scale parameter, alpha [25]. First, a network estimates the contour of the lung; then, the alpha-shape method smooths that boundary using the detected edge points. A "global mask" concept is applied to enforce anatomically realistic shapes, even in cases where disease partly obscures the true lung edge.

Testing this boundary-plus-alpha-shape approach showed a 4% gain in the Jaccard metric and a 2% gain in the Dice metric compared with a standard full-mask segmentation baseline [25]. These improvements

matter most in cases involving irregular lung shapes or where pleural effusion and consolidation distort the expected outline. The method also proved more resilient to differences in anatomy between patients, without needing extra training data — a useful property when data is limited.

4. AI in Software Implementation

The implementation phase — writing, reviewing, and debugging code — is the domain where AI-assisted tools have achieved the most visible commercial traction. LLMs for code generation represent one of the most active areas of software engineering research, with dozens of specialised models and evaluation benchmarks emerging in recent years.

3.3 Deep Learning Segmentation on Standard Datasets

Widely used lung-segmentation benchmarks include the Japanese Society of Radiological Technology (JSRT) dataset, the Shenzhen and Montgomery County tuberculosis datasets, and the NIH Chest X-ray14 dataset [26]. Performance is typically measured with the Dice similarity coefficient (DSC), the Jaccard index (IoU), Hausdorff distance, and sensitivity/specificity. Recent U-Net-based models report DSC scores above 0.98 on the JSRT dataset, suggesting near-perfect matches with human-drawn annotations for normal radiographs [27].

Cases involving disease remain harder to handle. Conditions such as consolidation, pleural effusion, and pneumothorax change the shape of the lung boundary in ways that models often under-segment. To address this, researchers have explored multi-task learning (predicting segmentation masks and disease labels together), curriculum learning (gradually introducing more difficult training examples), and test-time augmentation (to make predictions more stable) [28]. Generative adversarial networks (GANs) have also been used to create synthetic pathological lung images for augmentation purposes, meaningfully improving segmentation performance for underrepresented conditions [29].

4. Lung Disease Classification

4.1 Multi-Class Classification Using CNNs

Sorting chest X-rays into several disease categories is typically framed as either a multi-label or multi-class problem, depending on whether more than one condition can appear in the same image. Transfer learning from large pre-trained CNNs remains the most common approach. VGG-16 and VGG-19, built from deep stacks of small 3×3 convolutional filters, serve as reliable feature extractors capturing both texture and structure [10]. ResNet models, thanks to their residual connections, allow very deep networks to train without vanishing gradients, and ResNet-101 has performed competitively on respiratory sound classification tasks [11].

DenseNet-121 — where every layer has access to the outputs of all earlier layers — has shown especially strong results for chest disease classification and underlies CheXNet, a model that outperformed practicing radiologists at detecting pneumonia in the NIH ChestX-ray14 dataset [30]. EfficientNet, which scales network depth, width, and input resolution together in a balanced way, achieves a strong trade-off between accuracy and computational cost, and has been applied to detecting COVID-19 from chest X-rays [31].

4.2 AI-Assisted Code Review
Automated code review is a complementary application of LLMs in implementation. Systems such as CORE (Code Quality Issue Resolution) leverage LLMs to identify and resolve technical debt, code smells, and anti-patterns [29]. LLM-based code review tools can explain the rationale behind suggested changes in natural language, improving developer understanding and buy-in — a capability absent from purely static analysis tools [30].

4.2 Detecting Pneumonia and COVID-19

Pneumonia is among the most frequently diagnosed findings on chest radiographs. Models trained on public pneumonia CXR collections have achieved sensitivity and specificity figures above 90% [32]. During the COVID-19 pandemic, the pressure on healthcare systems made rapid automated screening especially valuable. CNNs fine-tuned on COVID-19 CXR image collections showed strong discriminative power, though thorough cross-dataset testing was

necessary to avoid misleadingly high results caused by data leakage [33].

Comparative studies of CNN-based COVID-19 detectors reported accuracy ranging from 87% to 98%, depending on the dataset used, the augmentation strategy applied, and the chosen architecture [34]. Grad-CAM visualizations showed that these models tended to focus on bilateral ground-glass opacities and areas of consolidation — findings consistent with known COVID-19 pathology and lending clinical credibility to the models' decisions. Still, biases linked to patient demographics and differences between imaging sites in the training data point to a continued need for validation on independent patient groups [35].

4.3 Vision Transformers for Classifying Respiratory Conditions

Vision Transformers have been applied to chest X-ray classification partly to overcome a known weakness of CNNs — their built-in assumptions about locality and translation invariance can limit their ability to capture broader spatial patterns relevant to diffuse lung disease. ViTs pre-trained on large natural-image datasets and then fine-tuned on CXR data have delivered competitive results, and their self-attention maps offer a way to visualize which regions of the image most influenced a prediction [7].

For respiratory sound classification, spectrogram images generated from lung auscultation recordings act as inputs to ViT models, linking sound-based and image-based analysis by exploiting the ViT's strength at modeling patterns across both time and frequency. Using appropriate normalization (RMS versus EBU), suitable augmentation, and well-configured patch embeddings, a ViT model outperformed ResNet-101, VGG-16, and VGG-19 on a five-category respiratory sound classification task, reaching 91.04% accuracy [11].

4.4 Multimodal Vision-Language Models

Multimodal vision-language models (VLMs) pair an image encoder with a language model, enabling reasoning that draws on both images and text together. In chest X-ray research, systems such as CheXzero and CXR-LLaVA make use of paired image-report

data to build meaningful representations without needing manually assigned disease labels [36].

A comparison across four major benchmark datasets — MIMIC-CXR, CheXpert, NIH-14, and PadChest — found that pretrained VLMs perform particularly well in zero-shot conditions, especially on data that differs from their training distribution, suggesting they generalize semantically quite well [36]. However, fine-tuning them on cross-lingual data or datasets with noisy labels tended to hurt performance, indicating some vulnerability to shifts in domain. Standard supervised models that were not multimodal, by contrast, benefited considerably from fine-tuning on data matching their original domain. This suggests a hybrid strategy — combining broad multimodal pretraining with more targeted fine-tuning — may offer the best balance between generalization and domain-specific accuracy.

5. Datasets and Evaluation Protocols

Reliable benchmark datasets underpin reproducible research on lung image analysis. NIH ChestX-ray14 includes more than 112,000 frontal chest images from roughly 30,000 patients, with 14 disease labels derived through natural language processing of radiology reports [26]. CheXpert offers 224,316 radiographs from 65,240 patients and introduced an uncertainty-labeling scheme that has spurred research into learning from imperfect labels [37]. MIMIC-CXR links images with reports and clinical notes from Beth Israel Deaconess Medical Center, supporting multimodal research [38]. PadChest, compiled in Spanish, contains over 160,000 images from about 67,000 patients covering more than 170 radiographic findings, making it useful for studying cross-lingual generalization [36].

For segmentation research, the JSRT dataset provides 247 radiographs with manually labeled lung and cardiac boundaries, while the Montgomery County and Shenzhen datasets focus specifically on tuberculosis screening. Though heavily used, these datasets are comparatively small and may not reflect the full range of presentations seen in real clinical settings. Evaluation typically relies on DSC, IoU, and precision-recall curves for segmentation, while classification work is usually assessed using ROC-AUC, accuracy, sensitivity, specificity, and F1 score [27].

One important pitfall in evaluating these models is data leakage — when images from the same patient end up in both the training and test sets. Proper stratification by patient is essential for trustworthy performance figures. Increasingly, researchers are also validating models across multiple sites and scanner types to gauge how well they generalize to new acquisition settings and patient populations [35].

6. Federated Learning and Privacy-Preserving Approaches

Medical imaging data is highly sensitive, and regulations such as HIPAA and GDPR place tight restrictions on sharing it between institutions. Federated learning (FL) offers a workaround: multiple institutions can jointly train a shared model without ever exchanging raw patient data [39]. Under FL, only local model updates — gradients or weight changes — are sent to a central server for aggregation, most commonly using the Federated Averaging (FedAvg) algorithm, while the underlying data never leaves each participant's own environment.

In an image classification experiment using FedAvg on Fashion-MNIST (60,000 training images and 10,000 test images spanning 10 clothing categories), the federated approach reached 87.27% accuracy, comparable to what centralized training achieves [40]. This 28×28 grayscale classification task demonstrated that complex pattern recognition can be learned effectively even when training is distributed, supporting FL as a practical option in privacy-sensitive scenarios. Although some performance gap with centralized training still exists in certain contexts, continued work on efficient communication protocols, differential privacy, and secure multi-party computation is steadily closing that gap.

When applied to chest X-ray analysis, FL allows hospital networks to jointly train disease classifiers across varied patient populations while keeping patient data confidential. Research shows that FL models trained on diverse, multi-site CXR data can outperform models trained locally at a single institution, especially for rarer conditions where no individual site has enough examples on its own [39]. Combining FL with differential privacy adds formal privacy guarantees, though this typically comes at some cost to accuracy, requiring careful tuning of the privacy budget.

7. Cross-Domain Perspectives

Although it might seem unrelated to lung imaging, handwritten character recognition shares core architectural principles with medical image classification. Systems built to recognize handwritten Arabic script — used, in various forms, across more than 25 languages including Arabic, Persian, Kurdish, and Urdu — rely on multi-layer neural networks trained via backpropagation, along with specialized feature-extraction methods designed to distinguish characters that look nearly identical.

These cross-domain parallels are instructive: regularization strategies, architectural choices, and training procedures developed for one type of visual recognition task frequently carry over to entirely different tasks. The success of backpropagation-trained multilayer networks on high-dimensional pattern recognition, first proven in character recognition, helped lay the groundwork for the deep CNN architectures now used in lung image analysis.

8. Challenges and Future Research Directions

Despite significant advances, several obstacles still stand between current research and real clinical deployment. First, dataset bias is a serious concern — most large CXR datasets originate from hospitals in North America or Europe, which can limit how well models perform on populations with different disease

patterns, body types, and equipment. Prospective studies testing AI models across a wider range of demographic groups are badly needed [35].

Second, models must be interpretable and trustworthy before clinicians will adopt them widely. Techniques like Grad-CAM, SHAP, and integrated gradients generate visual explanations after the fact, but they don't fully expose the model's underlying reasoning. Building inherently interpretable architectures — such as prototype-based networks and concept bottleneck models — is a promising path toward earning physicians' trust [7].

Third, combining radiographic images with other data types — clinical notes, lab results, patient history — could meaningfully improve diagnostic accuracy. VLMs trained on paired image-report data represent early progress in this direction, but they still require thorough testing across varied datasets and languages [36]. Fourth, deploying these models in low-resource environments calls for lightweight architectures suited to edge devices, along with efficient quantization and compression methods. Finally, continual learning — allowing models to adapt to new data distributions without losing previously learned knowledge — is essential for keeping performance stable as imaging technology and patient populations change over time.

9. Summary of Reviewed Studies

Study	Task	Method	Dataset	Key Metric
Boundary-based segmentation [25]	Segmentation	Alpha-shape + semantic boundary	CXR (private)	Jaccard +4%, Dice +2%
ViT respiratory sound [11]	Classification	Vision Transformer + EBU norm	Lung audio spectrograms	Accuracy: 91.04%
CheXNet [30]	Classification	DenseNet-121	NIH ChestX-ray14	AUC: 0.914
CheXzero / CXR-LLaVA [36]	Classification	VLM zero-shot & fine-tuned	MIMIC, CheXpert, NIH, PadChest	Strong ZS, FT varies

FedAvg image classification [40]	Classification	Federated Averaging	Fashion-MNIST	Accuracy: 87.27%
COVID-19 CNN [34]	Classification	EfficientNet / ResNet	COVID-19 CXR	Accuracy: up to 98%
U-Net lung segmentation [22]	Segmentation	U-Net FCN	JSRT / Montgomery	DSC: >0.97
AI traffic management [19]	Detection (general)	CNN + ML pipeline	Simulation	Accuracy: 94.89%, F1: 96.76%
SAR oil spill R-CNN [13]	Detection (general)	R-CNN + GLCM + DRLBP	SAR satellite images	Improved detection rate

Table 1 summarizes selected studies discussed in this review, listing their methodology, dataset, and key results.

CONCLUSION

This paper has reviewed a wide body of work on deep-learning-based lung segmentation and disease classification from chest X-rays, touching on CNNs, Vision Transformers, boundary-driven segmentation, multimodal vision-language models, and federated learning. Several distinct threads of research contribute to this picture: ViT-based classification of respiratory sounds [11], boundary-alpha lung segmentation [25], evaluation of VLMs for chest X-ray interpretation [36], federated approaches to image classification [40], and cross-domain benchmarking against ANN-based recognition. Collectively, this body of work reflects how quickly AI-assisted pulmonary diagnostics is maturing.

Several key conclusions emerge: (1) transformer-based models are narrowing the performance gap with CNNs while also offering better interpretability; (2) segmentation methods that incorporate geometric boundary constraints deliver clear metric improvements over standard full-mask approaches; (3) vision-language models generalize well in zero-shot settings but need careful fine-tuning when applied across languages or with noisy labels; (4) federated learning provides a workable route to training models across institutions while protecting patient privacy; and (5) the biggest remaining hurdles

for real-world use are dataset diversity, demographic representation, and prospective clinical validation.

Future research should focus on building larger, more demographically diverse, multimodal chest X-ray datasets; incorporating continual learning to handle shifting data distributions; and designing models whose decisions are interpretable in ways that align with how radiologists actually reason. Bringing these strands of research together offers real promise for building AI systems for respiratory care that are robust, fair, and trustworthy on a global scale.

REFERENCES

1. WHO, "The top 10 causes of death," World Health Organization, 2020.
2. A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.
3. P. Rajpurkar et al., "Deep learning for chest radiograph diagnosis," *PLOS Medicine*, vol. 15, no. 11, 2018.
4. X. Wang et al., "Hospital-scale chest X-ray database and benchmarks," in *Proc. CVPR*, 2017, pp. 2097–2106.
5. A. Iakovidis and A. Papageorgiou, "An intelligent system for automatic segmentation of medical images," *J. Med. Syst.*, vol. 35, pp. 1271–1282, 2011.

6. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in MICCAI, 2015, pp. 234–241.
7. A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. ICLR, 2021.
8. P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv:1711.05225, 2017.
9. T. Ai et al., "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 in China," *Radiology*, vol. 296, pp. E32–E40, 2020.
10. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. ICLR, 2015.
11. B. Aljaddouh, D. Malathi, and F. Alaswad, "Multimodal disease detection and classification using breath sounds and vision transformer for improved diagnosis," *Procedia Computer Science*, vol. 235, pp. 1436–1444, 2024.
12. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, no. 6, pp. 610–621, 1973.
13. J. D. D. Jayaseeli, D. Malathi, F. Alaswad, B. Aljaddouh, A. Bhardwaj, and T. Sahay, "Detection of oil-spill in satellite-based synthetic aperture radar (SAR) images using neural network," in Proc. 4th IEEE Global Conference for Advancement in Technology (GCAT), 2023, pp. 1–6.
14. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. NeurIPS, 2012.
15. J. Yosinski et al., "How transferable are features in deep neural networks?" in Proc. NeurIPS, 2014.
16. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. CVPR, 2016, pp. 770–778.
17. C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 60, 2019.
18. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
19. D. Malathi, F. Alaswad, B. Aljaddouh, L. Ranganayagi, and R. Sangeetha, "AI-Powered Traffic Management: Improving Congestion Detection and Signal Regulation," in 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI), IEEE, 2025, pp. 899–904.
20. A. Vaswani et al., "Attention is all you need," in Proc. NeurIPS, 2017.
21. H. Touvron et al., "Training data-efficient image transformers and distillation through attention," in Proc. ICML, 2021.
22. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science*, vol. 9351, pp. 234–241, 2015.
23. Z. Zhou et al., "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis*, 2018, pp. 3–11.
24. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. CVPR, 2015.
25. B. Aljaddouh, D. Malathi, and F. Alaswad, "Boundary-based X-Ray Lung Segmentation: A Novel Approach for Improving Accuracy and Efficiency," in 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), IEEE, 2023, pp. 154–160.
26. X. Wang et al., "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks," in Proc. CVPR, 2017.
27. S. Candemir et al., "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 577–590, 2014.
28. A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in Proc. NeurIPS, 2017.
29. I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
30. P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv:1711.05225, 2017.
31. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019.
32. D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, pp. 1122–1131, 2018.

33. [33] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 from chest X-ray images," *Sci. Rep.*, vol. 10, p. 19549, 2020.
34. A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease from chest X-ray image using CNN," *Pattern Recognit. Lett.*, vol. 141, pp. 152–158, 2021.
35. E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nat. Med.*, vol. 25, pp. 44–56, 2019.
36. B. Aljaddouh, D. Malathi, and F. Alaswad, "Multimodal vision-language models in chest x-ray analysis: a study of generalization, supervision, and robustness," *Biomedical Engineering Letters*, vol. 16, no. 2, pp. 517–537, 2026.
37. J. Irvin et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proc. AAAI*, 2019.
38. A. E. W. Johnson et al., "MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports," *Sci. Data*, vol. 6, p. 317, 2019.
39. B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in *Proc. AISTATS*, 2017.
40. J. D. D. Jayaseeli, D. Malathi, B. Aljaddouh, F. Alaswad, A. Shah, and D. Choudhary, "Image classification using federated averaging algorithm," in *Proc. 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2023, pp. 675–680.
41. B. Aljaddouh, D. Malathi, and F. Alaswad, "CXL-Net: Efficient lung image segmentation for CT and X-rays in low-resource clinics," *Procedia Computer Science*, vol. 283, pp. 856–865, 2026.
42. B. Aljaddouh and D. Malathi, "Trends of using machine learning for detection and classification of respiratory diseases: Investigation and analysis," *Materials Today: Proceedings*, vol. 62, pp. 4651–4658, 2022, doi: 10.1016/J.MATPR.2022.03.120
43. B. Aljaddouh and D. Malathi, "Enhancing boundary accuracy in semantic segmentation of chest X-ray images using Gaussian process regression," *Biomedical Physics & Engineering Express*, vol. 11, no. 4, p. 045005, 2025.

HOW TO CITE: Anwar Muta Mohammad*, Abirami J., Sangeetha Banerjee, Deep Learning For Lung Segmentation And Disease Classification In Chest X-Ray Imaging, *Int. J. Sci. R. Tech.*, 2026, 3 (8), 989-997. <https://doi.org/10.5281/zenodo.22111024>