Interdisciplinary Journal of Virtual Learning in Medical Sciences

Original Article

Deep Learning Model for Pneumothorax Detection in Chest Radiographs: A Multicenter Retrospective Cross-Sectional Study

Reza Ibrahimi[1], Mohammad Reza Azimi-Aval[2]*, Jalal Karger[3], Seyed Zia Hejripour[4], Behrang Rezvani Kakhki[1], Reza Gerami[2], Hojat Ebrahiminik[5]

[1]*Department of Emergency Medicine, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*
[2]*Radiation Sciences Research Center, AJA University of Medical Sciences, Tehran, Iran*
[3]*Department of Radiology, Faculty of Medicine, AJA University of Medical Sciences, Tehran, Iran*
[4]*Department of Emergency Medicine, Be-sat Hospital, AJA University of Medical Sciences, Tehran, Iran*
[5]*Department of Interventional Radiology and Radiation Sciences Research Center, Aja University of Medical Sciences, Tehran, Iran*

ABSTRACT

Background: Pneumothorax is a common clinical condition characterized by the presence of air within the pleural space, occurring in about half of chest trauma cases. Its clinical presentation ranges from asymptomatic cases to severe conditions causing hemodynamic instability or death. Deep learning models offer transformative potential for both clinical diagnosis and medical education through automated detection and interactive training tools. This study sought to evaluate deep learning models for detecting pneumothorax in Chest Radiographs (CXRs), assessing their diagnostic accuracy and potential to enhance medical education.

Methods: This retrospective cross-sectional study was conducted between February 2022 and September 2023 to assess the performance of four deep learning models for pneumothorax detection: Mask Region-based Convolutional Neural Network (Mask R-CNN), Deep Labelling version 3 (DeepLabv3), You Only Look Once version 8 (YOLOv8), and the U-shaped CNN model (U-Net). The evaluation was conducted using 20,000 chest X-ray images sourced from three hospitals in Iran, along with three open source datasets, including PTX-498, PTX-227, and SIIM-ACR-Pneumothorax. Images were labeled by consensus from two radiologists and two traumatologists. Rather than applying a conventional percentage-based split, a tiered data strategy was applied: internal datasets for training and validation, and external datasets (CheXpert and NIH) for independent testing to verify generalizability. Each model was trained to detect pneumothorax by extracting features and performing segmentation. Performance was evaluated using sensitivity, specificity, precision, recall, and F1-score. The outputs were analyzed for integration into virtual learning platforms to train medical students in diagnosing pneumothorax.

Results: The YOLOv8 algorithm showed the best performance for detecting and localizing pneumothorax, achieving an F1 score of 0.68. The final model's precision was 0.79, and a recall of 0.60, and it worked best on chest X-ray images with 1024x1024 resolution, particularly showing greater accuracy in identifying larger pneumothoraces.

Conclusion: Integration of YOLOv8 into medical education has the potential to improve diagnostic training via interactive AI-based simulations. However, challenges remain in detecting smaller pneumothoraces, highlighting the need for further optimization.

Keywords: Pneumothorax, Algorithms, Artificial Intelligence, Trauma, Wounds and Injuries, Education, Deep Learning

*Corresponding author:
Mohammad Reza Azimi-Aval, Radiation Sciences Research Center, AJA University of Medical Sciences, Tehran, Iran.
Tel: +98 9125393785
Fax: +98 21 33216525
Email: Dr.mr.azimi@gmail.com

Please cite this paper as:
Ibrahimi R, Azimi-Aval MR, Karger J, Hejripour SZ, Rezvani Kakhki B, Gerami R, Ebrahiminik H. Deep Learning Model for Pneumothorax Detection in Chest Radiographs: A Multicenter Retrospective Cross-Sectional Study. Interdiscip J Virtual Learn Med Sci. 2025;16(3):2-13. doi: 10.30476/ijvlms.2025.106465.1333.

Received: 25-03-2025
Revised: 29-08-2025
Accepted: 30-08-2025

Introduction

Pneumothorax, defined as the accumulation of air between the visceral and parietal pleura, can result in the separation of the lung from the thoracic wall. This condition is of significant clinical concern worldwide, presenting with a broad spectrum of symptoms ranging from asymptomatic cases to severe respiratory distress and hemodynamic instability (1). Pneumothorax is typically classified as either traumatic or spontaneous, with chest pain and dyspnea being the most common symptoms (2). Diagnosis is commonly made through chest X-ray, ultrasound, Computed Tomography (CT) scan, or clinical observation, particularly after air extraction and symptom relief from needle thoracostomy (3).

While pneumothorax can be life-threatening, especially in trauma patients, its timely diagnosis is critical. Chest trauma is seen in approximately 60% of multiple trauma patients, and pneumothorax occurs in about half of these individuals, contributing to a 10% mortality rate (4, 5). Early identification is challenging, as chest X-ray interpretation can be inconsistent among clinicians due to the complex anatomy and subtle signs of pneumothorax (6). Moreover, about 20% of pneumothoraces are missed on chest X-ray, and occult pneumothorax, visible on CT but not on X-ray, occurs in 30% of severe chest trauma cases (7, 8). This situation increases the dependence on CT imaging, raising concerns about cost, radiation exposure, and delayed diagnosis in emergency settings (9, 10).

Artificial intelligence (AI) offers transformative potential in medical education by providing virtual learning environments where students can practice diagnostic skills. Deep learning models, such as Convolutional Neural Networks (CNNs), can simulate real-world diagnostic scenarios, enhancing training for medical students and residents. Recent advancements in AI have shown promising results in Computer-Aided Diagnosis (CAD), particularly in radiology. The CNNs have contributed significantly to improving diagnostic accuracy in cardiothoracic imaging, a field that includes chest X-rays, which are among the most commonly performed radiologic procedures (6). However, several technical challenges persist in developing AI models for pneumothorax detection. The appearance of pneumothorax on chest X-rays can be subtle and varies greatly, which makes accurate segmentation quite challenging (11). Chest radiographs often contain noise from scattered radiation, patient movement, or medical devices (such as tubes and lines), which can either mimic or obscure the features of pneumothorax (12). Additionally, pneumothorax cases are relatively rare compared to standard chest X-rays, leading to class imbalances that can negatively impact model performance (13). Despite these challenges, AI models have demonstrated performance comparable to that of specialists in identifying various thoracic conditions, such as pulmonary nodules, pneumonia, and tuberculosis (14, 15). This automated analysis capability addresses key limitations of traditional diagnostic methods by reducing reliance on subjective expert interpretation and minimizing resource-intensive manual procedures. Recent research has concentrated on various models tailored to different medical urgencies and emergencies. Consistent with our findings, some studies reported that the YOLO model had high accuracy in detecting pathologies in chest X-rays (16, 17). In medical education, these models can be integrated into interactive platforms, allowing learners to analyze Chest Radiographs (CXRs), receive real-time feedback, and improve diagnostic proficiency.

Nevertheless, many existing studies rely on open-source datasets and lack true external validation, limiting the real-world applicability of their models. For instance, a study by Kitamura and colleagues reported significant performance degradation when a model was tested on an external dataset (18). Furthermore, many AI models do not include segmentation of the pathology, which is crucial for localizing pneumothorax (19). Some models, such as Lunit Insight, have been commercially developed, but their datasets are

often small and geographically limited, which can hinder their generalizability (19). This study addresses these gaps by developing an AI model for pneumothorax detection using diverse datasets and rigorous validation, while exploring its application in medical education to train future clinicians through virtual simulations. The model was trained using a diverse set of images from both native and open source datasets to enhance generalizability, and evaluated through both internal and external validation to ensure reliability and segmentation capabilities.

Methods

Study Design and Setting

This retrospective cross-sectional study analyzed a dataset consisting of approximately 20,000 CXRs collected between February 4, 2023, and September 20, 2024, to assess the performance of four deep learning models for pneumothorax detection. The majority of the images were gathered from trauma patients who were referred to three hospitals: *Besat*, *Imam Reza*, and *Hasheminejad*, located in the cities of Mashhad and Tehran, IRAN. Additional images were sourced from publicly available datasets, including PTX-498, PTX-227, and SIIM-ACR-Pneumothorax.

Participants and Sampling

The sample size was determined based on the theoretical Vapnik–Chervonenkis (VC) dimension of the CNN model, as described by D'Souza and colleagues (20). The inclusion criteria encompassed chest radiographs from trauma patients aged 18 years and older with confirmed pneumothorax (PTX) diagnoses. Exclusion criteria included incomplete imaging data and unclear CXRs. Control images were randomly selected at a 4:1 ratio to ensure balance between PTX-positive and PTX-negative samples. All images were reviewed and labeled by a consensus of two experienced radiologists and two traumatologists. To ensure consistency across model comparisons, all images were resized to three standard resolutions: 640×640, 840×840, and 1024×1024 pixels. Rather than applying a conventional 70/15/15 split for training, validation, and testing, we implemented a tiered data strategy. Internal datasets—including PTX-227, PTX-498, SIIM-ACR, and images from *Besat*, *Imam Reza*, and *Hasheminejad* hospitals—were used for training and validation. External testing was exclusively performed on 300 PTX-positive cases from CheXpert and NIH datasets to assess generalizability across institutions, imaging protocols, and patient populations. This stratified division was consistently applied to all resolution-specific subsets to maintain comparable conditions for model development and evaluation. The tiered design was selected to mitigate dataset-specific biases and to simulate real-world deployment scenarios, where models are often exposed to data distributions different from those seen during training.

Tools / Instruments

The primary tool used in this study was a CNN-based deep learning algorithm. This architecture included four convolutional layers (kernel size=3×3, stride=1, ReLU activation) and three fully connected layers (512, 256, and 2 neurons with Softmax activation), designed to perform segmentation, feature extraction, and classification of PTX regions in CXRs. Batch normalization was applied after each convolutional layer. The study employed the following segmentation models:

U-Net

U-Net is a neural network architecture specifically designed for segmentation tasks and has been widely applied in chest radiography studies (6, 11). The term "U-Net" is derived from the architecture's design, which resembles the letter "U" when depicted visually. Its U-shaped encoder–decoder structure enables precise pixel-wise segmentation, which is particularly valuable in CXRs where pneumothorax boundaries are subtle and often linear or discontinuous. In our study, U-Net provided a benchmark for evaluating whether detailed pixel-level feature extraction could delineate the pleural

line despite the known limitations of portable CXR imaging, such as noise and overlapping anatomical structures.

DeepLabV3

Deep Labelling V3 (DeepLabV3), which is designed for semantic image segmentation using atrous convolution and multi-scale feature extraction, has demonstrated improved recognition of variable-sized regions in thoracic imaging. This model was particularly suitable for detecting pneumothorax, as PTX manifestations range from very small apical collections to large hemithorax collapses, and subtle findings may be obscured by ribs or devices. Incorporating DeepLabV3 enabled us to explore whether its multi-scale contextual learning improves sensitivity to these size and positional variations, which are frequently identified as challenges in previous studies.

Mask R-CNN

Mask Region-based Convolutional Neural Network (Mask R-CNN) provides both bounding-box detection and instance segmentation, addressing a common limitation in many earlier pneumothorax studies that were unable to localize the affected areas (19). In trauma cases, which often present multiple pneumothorax sites along with interfering elements like chest tubes and monitoring lines, Mask R-CNN facilitates simultaneous localization and segmentation even in complex imaging scenarios, making it well-suited for addressing the practical challenges encountered in CXR interpretation.

YOLOv8

You Only Look Once version 8 (YOLOv8) is a real-time object detection system. YOLO models have been repeatedly reported in recent studies to achieve high accuracy in detecting thoracic pathologies in CXRs while maintaining real-time performance (16, 17). We selected YOLO for its specific optimization in both speed and accuracy. This choice aligns with the clinical fact that pneumothorax is a time-sensitive condition where rapid detection is critical, as well as with our study's aim of integrating AI into medical education platforms for interactive learning.

Data Collection

The data collection process involved retrospective review and annotation of CXRs by radiologists and traumatologists. The datasets used are described in detail as follows:

- ***PTX-498 Dataset:*** Initially contained 498 images; 87 false-positive images were excluded after a radiologist review.
- ***PTX-227 Dataset:*** Initially contained 227 images; 18 false-positive images were excluded after review.
- ***SIIM-ACR-Pneumothorax:*** This dataset contained 12,047 CXRs and mask images, including 3,576 PTX cases. Labels were initially generated using Natural Language Processing (NLP) and validated by two radiologists. Ultimately, a total of 2,156 images were confirmed as true positives, while the remainder were excluded. False negatives were negligible during the review.
- ***Besat Hospital Dataset:*** A total of 3,000 images were retrospectively collected over six months from the trauma and emergency departments. An experienced traumatologist identified 158 PTX-positive cases, which were subsequently confirmed through consensus reviews by two radiologists and a pulmonologist.
- ***Imam Reza Hospital Dataset:*** A total of 4,000 images were prospectively collected over the course of a year and screened by an experienced radiologist. Since this was a general hospital, not limited to trauma patients, only 88 images were identified as positive for PTX.
- ***Hasheminejad Hospital Dataset:*** A total of 4,000 images were collected over six months from the emergency department. An experienced traumatologist identified 169 PTX-positive images, confirmed by consensus review.

Validity and Reliability

To assess generalizability, two publicly

available datasets—CheXpert and NIH—were reserved exclusively for external testing and were not used during model training or internal validation. From the CheXpert dataset, comprising over 190,000 anteroposterior CXRs, 150 PTX-positive images were randomly selected through visual inspection. Similarly, the NIH dataset, containing 120,000 CXRs depicting various pathologies, provided 150 PTX-positive images selected through visual assessment. Additionally, 800 negatively labeled images were randomly sampled as controls from both datasets.

The models were trained for 200 epochs with a patience value of 25 and optimized using the following loss functions:

• ***Segmentation:*** Binary Cross-Entropy.
• ***Bounding Box Localization:*** Complete Intersection over Union (CIoU).
• ***Classification:*** Categorical Cross-Entropy.

To maximize the reliability, all images were annotated via consensus by two experienced radiologists and two traumatologists. This multi-expert labeling reduced inter-observer variability and ensured consistent identification and delineation of pneumothorax regions. Cases with unclear or incomplete imaging were excluded to maintain data quality.

Prior to analysis, the data underwent preprocessing to improve image quality and enhance the robustness of the model. This process involved normalization, noise reduction using MedianBlur from Albumentations, and various data augmentation techniques such as rotation, horizontal flipping, contrast and brightness adjustments, gamma correction, elastic transformations, grid distortions, and optical distortions to simulate real-world variability.

Data Analysis

The dataset was split into training, validation, and testing sets to optimize the performance and generalizability of the deep learning models. It was comprised of a total of 3,191 cases, and 12,764 control images. The training dataset included 2,934 images distributed as follows: 209 cases from PTX-227, 411 cases from PTX-498, 2,156 cases from SIIM-ACR-Pneumothorax, and 158 cases from the *Besat* Hospital dataset. A total of 257 cases were used to validate and fine tune the model, in which *Imam Reza* and *Hashemi Nejad* hospitals contributed 88 and 169 cases respectively. As for the test set, external testing was performed on 300 PTX-positive cases equally selected from Chexpert and NIH datasets by visual inspection. During training, class imbalance was addressed through weighted loss functions (positive class weight=4) and balanced batch sampling, while preserving the natural 4:1 distribution in test sets. Additionally, performance metrics, including sensitivity, specificity, precision, recall, and F1-score, were calculated to evaluate the models' diagnostic accuracy. All analyses were performed using Python 3.9 on Ubuntu 20.04 LTS, employing PyTorch 1.10.0 with torchvision 0.11.0 for U-Net, DeepLabv3, and Mask R-CNN architectures, the Ultralytics YOLOv8 8.2.93 package, Albumentations 1.1.0 for data augmentation, and an NVIDIA RTX 3090 GPU with 24GB VRAM accelerated by CUDA 11.7.

Ethics - The study received ethical approval from the Ethics Committee of AJA University of Medical Sciences, Tehran, Iran. Informed consent was waived due to the retrospective nature of the study and the use of anonymized data. All data were handled in accordance with relevant ethical guidelines to ensure patient confidentiality and data security. All patient information was anonymized at the point of collection by removing sensitive data, including names, IDs, and examination dates.

Results

The performance of the segmentation models, including U-Net, DeepLabv3, Mask R-CNN, and YOLOv8, was evaluated using various image resolutions and confidence scores. As shown in Table 1 and Figure 1, the U-Net model was trained with a resolution of 840×840 pixels. It demonstrated limited performance in segmenting PTX regions, with low recall and moderate precision.

Table 1: Performance metrics for different segmentation models across various image resolutions

Model	Image size	Precision	Recall	F1-score
U-Net	840×840	0.51	0.27	0.35
DeepLabV3	840×840	0.53	0.26	0.36
Mask R-CNN	840×840	0.56	0.32	0.40
YOLOv8	640×640	0.55	0.32	0.40
	840×840	0.63	0.44	0.52
	1024×1024	0.79	0.60	0.68

DeepLabv3: Deep Labelling Version 3; Mask R-CNN: Mask Region-based Convolutional Neural Network; YOLOv8: You Only Look Once version 8; U-Net: U-Shaped Convolutional Neural Network

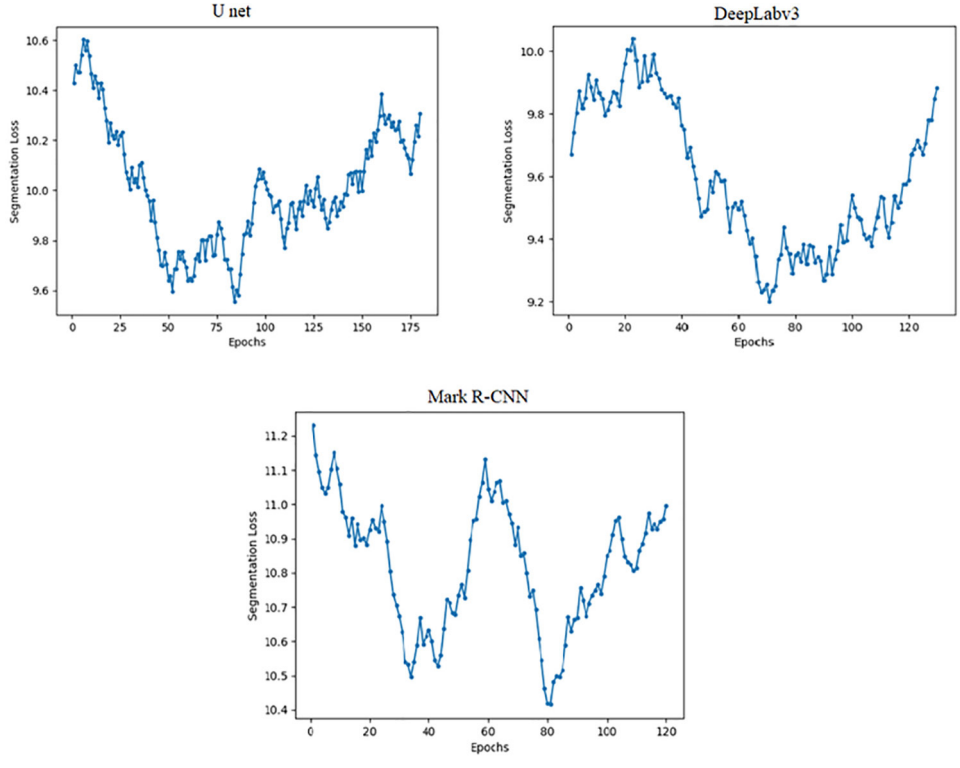

Figure 1: Segmentation loss function for U-Net, DeepLabv3, and Mask R-CNN. DeepLabv3: Deep Labelling Version 3; Mask R-CNN: Mask Region-based Convolutional Neural Network; U-Net: U-Shaped Convolutional Neural Network.

The DeepLabv3 model achieved slightly better precision compared to U-Net but exhibited similar limitations in recall, indicating challenges in detecting all positive regions. The Mask R-CNN model demonstrated superior segmentation performance compared to U-Net and DeepLabv3, achieving the highest F1-score among the three. However, recall remained suboptimal, highlighting the need for further optimization. The YOLOv8 model was evaluated at three image resolutions (640×640, 840×840, and 1024×1024 pixels, with the best results obtained at 1024×1024 pixels, where it achieved the highest F1-score and precision across all tested configurations (Figure 2).

Various confidence thresholds were tested for the classification task to identify the optimal balance between precision and recall. The optimal probability score for classifying pneumothorax was empirically determined to be 0.03 after evaluating precision-recall trade-offs across multiple scores. At this score, the model achieved balanced performance (precision=0.82, recall=0.82, and F1=0.82). Higher thresholds increased precision but reduced recall, which is crucial for maintaining clinical sensitivity (Table 2).

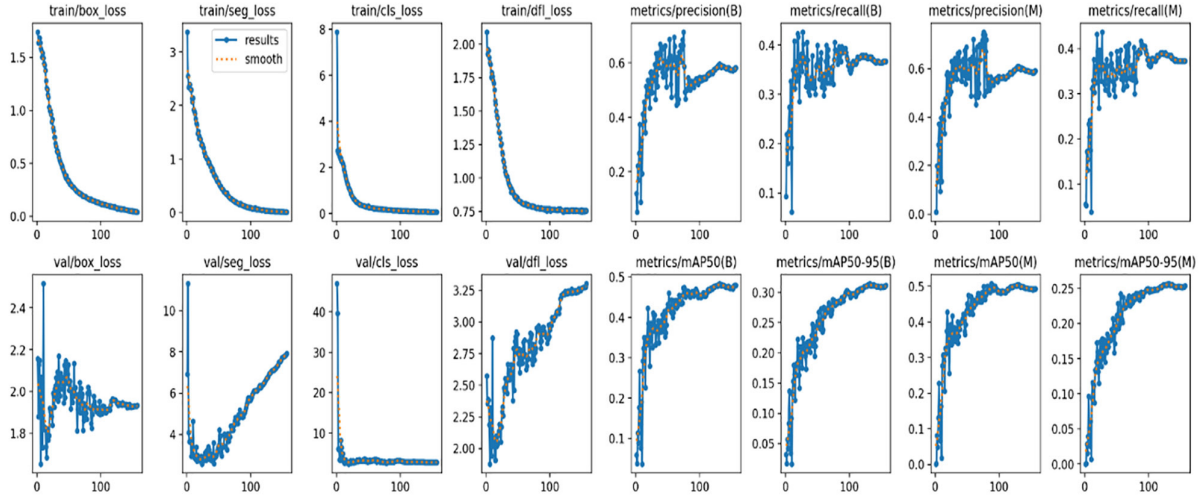

Figure 2: The segmentation loss and classification loss during the training and validation phases of the YOLOv8 model. *YOLOv8: You Only Look Once version 8.

Table 2: Performance metrics for classification across different confidence thresholds

Confidence Threshold	Precision	Recall	F1-score
0.30	0.94	0.39	0.55
0.23	0.91	0.44	0.60
0.10	0.81	0.56	0.66
0.08	0.79	0.59	0.68
0.05	0.76	0.66	0.70
0.03	0.82	0.82	0.82
0.01	0.61	0.82	0.70

The YOLOv8 model with a resolution of 1024×1024 pixels showed the best segmentation performance, achieving an F1-score of 0.68. The classification task achieved the highest overall performance at a confidence threshold of 0.03, with precision, recall, and F1-score all equal to 0.82. Moreover, a higher confidence threshold improved precision but significantly decreased recall, which is a critical metric for clinical applications.

Discussion

The results of this study demonstrated that YOLOv8 outperforms other prominent segmentation models, such as U-Net, DeepLabv3, and Mask R-CNN, in detecting pneumothorax. It achieved its highest F1-score of 0.68 at a resolution of 1024×1024 pixels, which represents a 70% to 94% improvement over competing architectures. Notably, the model demonstrated a balanced precision of 0.79 and a recall of 0.60. This addresses a significant issue in previous studies, where high precision often compromised the clinically important aspect of sensitivity.

Due to institutional and clinical resource constraints, efforts focused on the most effective outcome. Initially, a baseline comparison was established, where all models were compared at a uniform 840×840 resolution. At this benchmark, YOLOv8 (F1-score of 0.52) significantly outperformed all other models, thereby making it the most logical candidate for further evaluation. Higher resolutions like 1024×1024 demand significantly greater computational resources and can lead to memory bottlenecks and a reduction in usable batch size (20, 21). U-Net, DeepLabv3, and Mask R-CNN models all rely on more complex, multi-stage architectures that are inherently slower and more computationally taxing, making them ill-suited for high-resolution (22). In contrast, YOLOv8's single-stage architecture

is renowned for its efficiency and speed, allowing it to scale more effectively to higher resolutions without the same performance penalties. The exclusive evaluation of YOLOv8 at 1024×1024 warranted the dual benefit of superior accuracy and computational feasibility for general clinical use.

For classification, an optimal confidence threshold of 0.03 was identified, yielding a well-balanced precision, recall, and F1-score of 0.82 each. This balance is critical in practice, as low recall risks missing diagnoses. These findings underscore YOLOv8's potential as a fast, accurate, and deployable CAD solution for pneumothorax detection, especially due to its web-based, real-time inference capabilities.

Early detection is critical in managing time-sensitive medical conditions, and pneumothorax is no exception. Timely diagnosis and treatment of tension pneumothorax can significantly reduce mortality rates, which range from 3% to 7% when treated promptly but escalate to over 90% in cases of delayed or inappropriate management (23). Given the straightforward and cost-effective treatment options for pneumothorax, enhancing diagnostic efficiency is vital. CAD systems have emerged as promising tools, offering the potential to reduce diagnostic time and improve patient outcomes drastically (24). For instance, one study demonstrated a significant reduction in pneumothorax diagnosis time from 9.8±2 to 1±0.5 minutes using a CAD model (25).

Recent studies highlight the critical role of AI in medical education, demonstrating that it greatly enhances trainees' accuracy in diagnosing conditions across various imaging modalities. This includes interpretations of hip fracture X-rays, brain MRIs, clinical reasoning exercises, and enhanced surgical practices (26-29). Despite these advancements, many algorithms in the field have been developed for specific clinical scenarios, such as post-procedural pneumothorax detection.

For instance, Park and colleagues developed an algorithm using a dataset of 1,596 post-biopsy pneumothorax cases, achieving an AUC of 0.905. However, the specificity of this model to post-procedural pneumothorax limits its applicability to broader populations, including trauma-related cases (30). Aligning with our findings, Tseng WC and colleagues also demonstrated AI's strong potential to enhance diagnostic efficiency and accuracy, particularly in complex and high-volume pneumothorax detection on chest CXRs (31). Our study aimed to address this gap by developing a model trained on a diverse dataset with rigorous labeling standards. Unlike many other datasets used in CAD development, which often rely on pre-labeled cases from competitions, our dataset was annotated through a consensus among two radiologists and two traumatologists, ensuring a high standard of reliability (7). This approach enhanced the clinical relevance and generalizability of our findings.

External validation remains a critical challenge for CAD models. For instance, Kim and colleagues (32) reported a sensitivity of 90% for their pneumothorax model during internal validation, but this dropped to 69% when benchmarked against CT scans, the gold standard for diagnosis. This underscores the importance of rigorous external validation to assess the real-world utility of such models.

Additionally, the web-based, real-time serving capability of our YOLOv8 model enhanced its clinical applicability, particularly in controlled and monitored environments. The confidence interval used in deployment ensured reliable performance, making it a practical tool for integration into clinical workflows.

Other architectures, such as U-Net, DeepLabv3, and Mask R-CNN, showed varied performance, with specific strengths in segmentation and precision metrics. However, YOLOv8 outperformed these models in terms of overall accuracy and real-time diagnostic potential.

Moving forward, efforts should focus on improving the recall of the YOLOv8 model, particularly for smaller pneumothoraces, and exploring its integration into clinical practice

for broader validation. The findings of this study highlight the potential of CAD systems to augment traditional diagnostic approaches, ultimately contributing to better patient outcomes through faster and more accurate pneumothorax detection.

Limitations and Suggestions

This study faced several limitations. First, while our comparison of four advanced segmentation models provided robust benchmarks, the predominance of native datasets (79.8% of images) raises potential concerns about overfitting to institution-specific patterns and label noise from single-center annotations. Second, variations in image acquisition protocols and labeling methods across datasets may introduce biases, despite our rigorous validation process. Third, hardware constraints limited resolution standardization across models; for instance, YOLOv8 was evaluated at a resolution of 1024×1024, while the others used 840×840 resolution. Future works should incorporate multicenter external validation and consensus labeling to address these limitations.

Conclusion

The results of our study indicate that the YOLOv8-based model achieved a well-balanced performance for pneumothorax detection in trauma patients, with precision, recall, and F1-score all at 0.82 at the optimal threshold of 0.03. This model surpassed other architectures in segmentation tasks, which showed an F1-score of 0.68 at 1024×1024 resolution. Although these findings highlight the model's promise for supporting radiologists and emergency physicians in trauma care, additional validation is necessary to confirm its effectiveness and integration in real clinical workflows.

Abbreviations

AI: Artificial Intelligence
CAD: Computer-Aided Diagnosis
CIoU: Complete Intersection over Union
CNN: Convolutional Neural Network
CXR: Chest Radiograph
DeepLabv3: Deep Labelling Version 3
Mask R-CNN: Mask Region-based Convolutional Neural Network
NLP: Natural Language Processing
PTX: Pneumothorax
U-Net: A U-Shaped Convolutional Neural Network
YOLOv8: You Only Look Once Version 8

Acknowledgments

The authors express their gratitude to all radiologists and traumatologists who contributed to this study.

Authors' Contribution

RI, BRK, and MRAA were responsible for the initial drafting of the manuscript. JK and HE oversaw data gathering, and SZH, RG and RI performed the statistical analysis and validation. MRAA was responsible for revisions. All the authors approved the final manuscript.

Conflict of Interest

The authors declare that they have no conflicts of interest to disclose.

Ethical Considerations

The study was conducted following ethical guidelines and received approval from the Research Ethics Committee of the AJA University of Medical Sciences, Tehran, Iran (approval code IR.AJAUMS.REC.1402.219). Informed consent was waived due to the retrospective nature of the study and the use of anonymized data. The study design and execution adhered to the university's established rules and protocols.

Funding/Support

This study was not supported by any external funding sources.

Availability of Data and Materials

The detailed datasets and supplementary information that may enhance the interpretation of the research results will be accessible upon reasonable request from the corresponding author.

References

1 Swierzy M, Helmig M, Ismail M, Rückert J, Walles T, Neudecker J. [Pneumothorax]. Zentralbl Chir. 2014;139 Suppl 1:S69-86; quiz S7. German. doi: 10.1055/s-0034-1383029. PubMed PMID: 25264729.

2 DeMaio A, Semaan R. Management of Pneumothorax. Clin Chest Med. 2021;42(4):729-38.

3 Tran J, Haussner W, Shah K. Traumatic Pneumothorax: A Review of Current Diagnostic Practices And Evolving Management. J Emerg Med. 2021;61(5):517-28. PubMed PMID: 34470716.

4 Ince A, Ozucelik DN, Avci A, Nizam O, Dogan H, Topal MA. Management of pneumothorax in emergency medicine departments: multicenter trial. Iran Red Crescent Med J. 2013;15(12):e11586. PubMed PMID: 24693384; PubMed Central PMCID: PMC3955499.

5 Alghnam S, Aldahnim MH, Aldebasi MH, Towhari JA, Alghamdi AS, Alharbi AA, Almarhabi YA, Albabtain IT. The incidence and predictors of pneumothorax among trauma patients in Saudi Arabia. Findings from a level-I trauma center. Saudi Med J. 2020;41(3):247-52. doi: 10.15537/smj.2020.3.24917. PubMed PMID: 32114596; PubMed Central PMCID: PMC7841555.

6 Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: A survey. Med Image Anal. 2021;72:102125. doi: 10.1016/j.media.2021.102125. PubMed PMID: 34171622.

7 Thian YL, Ng D, Hallinan J, Jagmohan P, Sia SY, Tan CH, et al. Deep Learning Systems for Pneumothorax Detection on Chest Radiographs: A Multicenter External Validation Study. Radiol Artif Intell. 2021;3(4):e200190. doi: 10.1148/ryai.2021200190. PubMed PMID: 34350409; PubMed Central PMCID: PMC8328109.

8 Lee KL, Graham CA, Yeung JH, Ahuja AT, Rainer TH. Occult pneumothorax in Chinese patients with significant blunt chest trauma: incidence and management. Injury. 2010;41(5):492-4. doi: 10.1016/j.injury.2009.12.017. PubMed PMID: 20097342.

9 Tian Y, Wang J, Yang W, Wang J, Qian D. Deep multi-instance transfer learning for pneumothorax classification in chest X-ray images. Med Phys. 2022;49(1):231-43. doi: 10.1002/mp.15328. PubMed PMID: 34802144.

10 Li X, Thrall JH, Digumarthy SR, Kalra MK, Pandharipande PV, Zhang B, et al. Deep learning-enabled system for rapid pneumothorax screening on chest CT. Eur J Radiol. 2019;120:108692. doi: 10.1016/j.ejrad.2019.108692. PubMed PMID: 31585302.

11 Wang Q, Liu Q, Luo G, Liu Z, Huang J, Zhou Y, et al. Automated segmentation and diagnosis of pneumothorax on chest X-rays with fully convolutional multi-scale ScSE-DenseNet: a retrospective study. BMC Medical Informatics and Decision Making. 2020;20(14):317. doi: 10.1186/s12911-020-01325-5. PubMed PMID: 33323117; PubMed Central PMCID: PMC7739478.

12 Jensen L, Meyer C. Reducing errors in portable chest radiography. Applied Radiology. 2015;44(4):7-15.

13 Salam A, Andono PN, Purwanto, Soeleman MA, Sidiq M, Alzami F, et al. NCT-CXR: Enhancing Pulmonary Abnormality Segmentation on Chest X-Rays Using Improved Coordinate Geometric Transformations. J Imaging. 2025;11(6):186. doi: 10.3390/jimaging11060186. PubMed PMID: 40558785; PubMed Central PMCID: PMC12194474.

14 Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. Eur J Radiol. 2020;123:108774. doi: 10.1016/j.ejrad.2019.108774. PubMed PMID: 31841881.

15 Do S, Song KD, Chung JW. Basics of Deep Learning: A Radiologist's Guide to Understanding Published Radiology

Articles on Deep Learning. Korean J Radiol. 2020;21(1):33-41. doi: 10.3348/kjr.2019.0312. PubMed PMID: 31920027; PubMed Central PMCID: PMC6960318.

16 Xie Y, Zhu B, Jiang Y, Zhao B, Yu H. Diagnosis of pneumonia from chest X-ray images using YOLO deep learning. Front Neurorobot. 2025;19:1576438. doi: 10.3389/fnbot.2025.1576438. PubMed PMID: 40370634; PubMed Central PMCID: PMC12077197.

17 Hao S, Li X, Peng W, Fan Z, Ji Z, Ganchev I. YOLO-CXR: A novel detection network for locating multiple small lesions in chest X-ray images. IEEE Access. 2024;12:156003-19. doi: 10.1109/ACCESS.2024.3482102.

18 Kitamura G, Deible C. Retraining an open-source pneumothorax detecting machine learning algorithm for improved performance to medical images. Clin Imaging. 2020;61:15-9. doi: 10.1016/j.clinimag.2020.01.008. PubMed PMID: 31954346; PubMed Central PMCID: PMC7085967.

19 Hwang EJ, Hong JH, Lee KH, Kim JI, Nam JG, Kim DS, et al. Deep learning algorithm for surveillance of pneumothorax after lung biopsy: a multicenter diagnostic cohort study. Eur Radiol. 2020;30(7):3660-71. doi: 10.1007/s00330-020-06771-3. PubMed PMID: 32162001.

20 D'Souza R N, Huang PY, Yeh FC. Structural Analysis and Optimization of Convolutional Neural Networks with a Small Sample Size. Sci Rep. 2020;10(1):834. doi: 10.1038/s41598-020-57866-2. PubMed PMID: 31965034; PubMed Central PMCID: PMC6972775.

21 Sabottke CF, Spieler BM. The Effect of Image Resolution on Deep Learning in Radiography. Radiol Artif Intell. 2020;2(1):e190015. doi: 10.1148/ryai.2019190015. PubMed PMID: 33937810; Pubmed Central PMCID: PMC8017385.

22 Suhaimi MS, Zainuddin K, Ghazali MD, Marzukhi F, Samad AM, Majid Z, et al., editors. Comparison of One-Stage and Two-Stage Strategies of Machine Learning Model for Rock Art Object Detection. 2023 IEEE 13th International Conference on System Engineering and Technology (ICSET); 2023 Oct 2; Shah Alam, Selangor, Malaysia. USA: IEEE Xplore; 2023. P. 215-20. doi: 10.1109/ICSET59111.2023.10295089.

23 Hoechter DJ, Speck E, Siegl D, Laven H, Zwissler B, Kammerer T. Tension Pneumothorax During One-Lung Ventilation - An Underestimated Complication? J Cardiothorac Vasc Anesth. 2018;32(3):1398-402. doi: 10.1053/j.jvca.2017.07.022. PubMed PMID: 29361455.

24 Zarogoulidis P, Kioumis I, Pitsiou G, Porpodis K, Lampaki S, Papaiwannou A, et al. Pneumothorax: from definition to diagnosis and treatment. J Thorac Dis. 2014;6(Suppl 4):S372-6. doi: 10.3978/j.issn.2072-1439.2014.09.24. PubMed PMID: 25337391; PubMed Central PMCID: PMC4203989.

25 Feng S, Liu Q, Patel A, Bazai SU, Jin CK, Kim JS, et al. Automated pneumothorax triaging in chest X-rays in the New Zealand population using deep-learning algorithms. J Med Imaging Radiat Oncol. 2022;66(8):1035-43. doi: 10.1111/1754-9485.13393. PubMed PMID: 35224858.

26 Rudie JD, Duda J, Duong MT, Chen P-H, Xie L, Kurtz R, et al. Brain mri deep learning and bayesian inference system augments radiology resident performance. J Digit Imaging. 2021;34(4):1049-58. doi: 10.1007/s10278-021-00470-1. PubMed PMID: 34131794; PubMed Central PMCID: PMC8455800.

27 Cheng CT, Chen CC, Fu CY, Chaou CH, Wu YT, Hsu CP, Chang CC, Chung IF, Hsieh CH, Hsieh MJ, Liao CH. Artificial intelligence-based education assists medical students' interpretation of hip fracture. Insights Imaging. 2020;11(1):119. doi: 10.1186/s13244-020-00932-0. PubMed PMID: 33226480; PubMed Central PMCID: PMC7683624.

28 Wang M, Sun Z, Jia M, Wang Y, Wang H, Zhu X, et al. Intelligent virtual case learning system based on real medical records and natural language processing. BMC Med Inform Decis Mak. 2022;22(1):60. doi: 10.1186/s12911-022-01797-7.

29 Blumstein G, Zukotynski B, Cevallos N, Ishmael C, Zoller S, Burke Z, et al. Randomized trial of a virtual reality tool to teach surgical technique for tibial shaft fracture intramedullary nailing. Journal of surgical education. 2020;77(4):969-77. doi: 10.1016/j.jsurg.2020.01.002. PubMed PMID: 32035854; PubMed Central PMCID: PMC7351249.

30 Park S, Lee SM, Kim N, Choe J, Cho Y, Do KH, Seo JB. Application of deep learning-based computer-aided detection system: detecting pneumothorax on chest radiograph after biopsy. Eur Radiol. 2019;29(10):5341-8. doi: 10.1007/s00330-019-06130-x. PubMed PMID: 30915557.

31 Tseng WC, Wang YC, Chen WC, Lin KP. Development of an AI model for pneumothorax imaging: Dataset and model optimization strategies for real-world deployment. Eur J Radiol Open. 2025;14:100664. doi: 10.1016/j.ejro.2025.100664. PubMed PMID: 40547323; PubMed Central PMCID: PMC12179720.

32 Kim D, Lee JH, Kim SW, Hong JM, Kim SJ, Song M, Choi JM, Lee SY, Yoon H, Yoo JY. Quantitative Measurement of Pneumothorax Using Artificial Intelligence Management Model and Clinical Application. Diagnostics (Basel). 2022;12(8). doi: 10.3390/diagnostics12081823. PubMed PMID: 36010174; PubMed Central PMCID: PMC9406694.