Check for updates

Received: 20 December 2023 | Accepted: 4 June 2024

DOI: 10.1002/ppul.27133

ORIGINAL ARTICLE

WILEY

XRAInet: AI-based decision support for pneumothorax and pleural effusion management

Mustafa Alper Akay MD[1] | Ozan Can Tatar MD[2] | Elif Tatar MD[1] | Beyza Nur Tağman MD[1] | Semih Metin MD[1] | Onursal Varlıklı MD[1]

[1]Department of Pediatric Surgery, Faculty of Medicine, Kocaeli University, Kocaeli, Turkey

[2]Department of General Surgery, Faculty of Medicine, Kocaeli University, Kocaeli, Turkey

Correspondence
Ozan Can Tatar, MD, Department of General Surgery, Faculty of Medicine, Kocaeli University, Kocaeli 41000, Turkey.
Email: ozancan.tatar@kocaeli.edu.tr

Funding information
None

Abstract

Purpose: This study aimed to develop and assess the performance of an artificial intelligence (AI)-driven decision support system, XRAInet, in accurately identifying pediatric patients with pleural effusion or pneumothorax and determining whether tube thoracostomy intervention is warranted.

Methods: In this diagnostic accuracy study, we retrospectively analyzed a data set containing 510 X-ray images from 170 pediatric patients admitted between 2005 and 2022. Patients were categorized into two groups: Tube (requiring tube thoracostomy) and Conservative (managed conservatively). XRAInet, a deep learning-based algorithm, was trained using this data set. We evaluated its performance using various metrics, including mean Average Precision (mAP), recall, precision, and F1 score.

Results: XRAInet, achieved a mAP score of 0.918. This result underscores its ability to accurately identify and localize regions necessitating tube thoracostomy for pediatric patients with pneumothorax and pleural effusion. In an independent testing data set, the model exhibited a sensitivity of 64.00% and specificity of 96.15%.

Conclusion: In conclusion, XRAInet presents a promising solution for improving the detection and decision-making process for cases of pneumothorax and pleural effusion in pediatric patients using X-ray images. These findings contribute to the expanding field of AI-driven medical imaging, with potential applications for enhancing patient outcomes. Future research endeavors should explore hybrid models, enhance interpretability, address data quality issues, and align with regulatory requirements to ensure the safe and effective deployment of XRAInet in healthcare settings.

KEYWORDS
decision support systems, deep learning, diagnostic imaging, pleural effusion, pneumothorax

1 | INTRODUCTION

Pneumothorax and pleural effusion are significant medical conditions in pediatric patient population that require prompt diagnosis and treatment to prevent severe consequences.[1,2] Accurately identifying and decision making, especially with X-ray imaging, is challenging, particularly for nonsurgical and primary care teams with limited experience, working in a low-resource setting.[3]

Deep learning (DL) techniques, particularly object detection algorithms, have emerged as powerful tools for automating the

 © 2024 Wiley Periodicals LLC.

detection and localization of abnormalities in medical images.[4,5] These algorithms train neural networks to classify and locate specific objects within images, offering potential assistance in identifying pneumothorax and pleural effusion regions in chest X-ray images.[6]

While recent studies have explored the use of DL algorithms for detecting pneumothorax and pleural effusion, accurately determining the need for surgical intervention remains a challenge. The decision to perform tube thoracostomy requires a comprehensive evaluation that incorporates clinical expertise, surgical experience, and diagnostic algorithms.[7]

To address these challenges, we have developed XRAInet, an innovative artificial intelligence (AI)-driven decision support system that aims to enhance the accuracy and efficiency of diagnosing critical medical conditions, including pneumothorax and pleural effusion. By integrating surgical experience, diagnostic algorithms, clinical intuition, and accumulated knowledge into the DL model, XRAInet aims to improve decision-making processes and ultimately enhance patient outcomes.

One notable advantage of XRAInet is its utilization of the YOLO DL algorithm as a baseline model, known for its efficiency and real-time universal object detection capabilities.[8] This algorithm enables fast processing of images and accurate identification and localization of the region of interest.

Furthermore, the widespread availability and cost-effectiveness of X-ray imaging make it a suitable modality for developing automated models like XRAInet. Despite its spatial resolution limitations compared to CT scans, X-ray imaging remains widely implemented in various healthcare settings, making XRAInet a potential practical solution for improving diagnostic capabilities.[9]

Our study focuses on evaluating the effectiveness of XRAInet in detecting and providing decision support for cases of pneumothorax and pleural effusion using X-ray images of those who required intervention. This study contributes to the field of AI-driven decision support systems in medical imaging, specifically in the management of pneumothorax and pleural effusion. Our findings will provide insights into the potential of DL techniques in enhancing diagnostic accuracy and improving patient outcomes for critical medical conditions. Furthermore, this research may guide future developments, including the exploration of hybrid models that integrate both X-ray and CT imaging for more comprehensive assessments.

2 | MATERIALS AND METHODS

2.1 | Patient selection and data collection

This study analyzed 250 patients admitted with pleural effusion and pneumothorax between 2005 and 2022. The patients were selected based on specific inclusion criteria, which involved having an X-ray at submission performed at our clinic, having a tube thoracostomy procedure for pleural effusion or pneumothorax or conservative management during the stay, and uneventful discharge at the end of the hospital stay. 170 patients were included in the study regarding the inclusion criteria. Patient distribution regarding exclusion criteria can be seen in Table 1.

2.2 | Patient grouping

The patients in this study were categorized into two distinct groups: the Tube (Tube Thoracostomy) group and the Conservative (Conservative Management) group.

2.2.1 | Tube group

Patients falling under the Tube group were further divided into two subgroups based on their diagnosis:

Tube on Admission: This subgroup comprised patients who received tube thoracostomy immediately upon admission. After their procedure, they were discharged home uneventfully or with Grades 1–2 Clavien–Dindo complications.

Delayed Tube: The Delayed Tube subgroup included patients who initially received a diagnosis of pleural effusion or pneumothorax but were managed nonoperatively for a period. However, their condition worsened during the nonoperative wait period, necessitating subsequent intervention.

After their procedure, they were discharged home uneventfully or with Grades 1–2 Clavien–Dindo complications.

2.2.2 | Conservative group

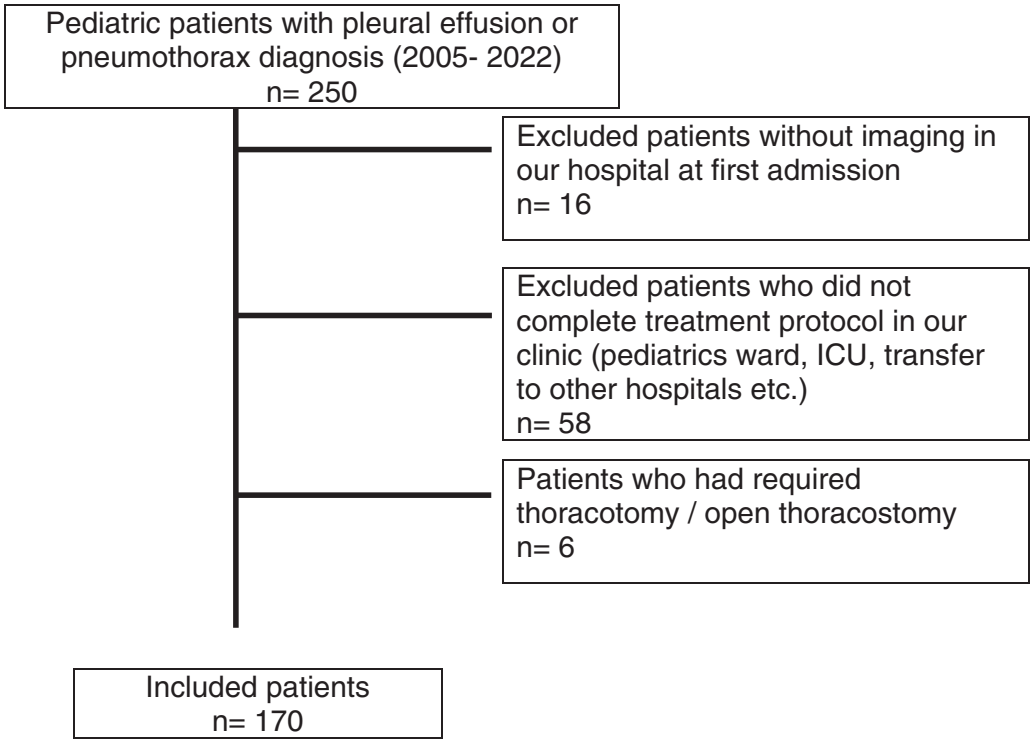
The conservative management group encompasses patients who did not require intervention and whose nonoperative management proceeded successfully. These patients initially presented with pleural effusion or pneumothorax findings confirmed by radiology; however, their conditions improved over time without the need for intervention. After their stay, they were discharged home uneventfully or with Grades 1–2 Clavien–Dindo complications. This group serves as a vital point of comparison, shedding light on the effectiveness of nonoperative approaches in managing cases that initially exhibited pleural effusion or pneumothorax-related findings.

2.3 | DL-based object detection algorithm (XRAInet)

XRAInet architecture was constructed using Python programming language (version 3.8) and TensorFlow (version 2.0.0), an open-source DL framework developed by Google. The YOLO algorithm was employed as the baseline model for XRAInet. YOLO is known for its efficiency and fast processing, making it suitable for real-time object detection tasks.

10990496, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppul.27133 by Mustafa Alper Akay - Kocaeli Üniversitesi , Wiley Online Library on [04/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Patient distribution regarding exclusion criteria.



The object detection model was trained using a data set of X-ray images from patients with pneumothorax and pleural effusion, along with annotations indicating the regions requiring tube thoracostomy. The training and validation of the model were performed on a computer system equipped with high-performance hardware specifications.

2.4 | Data annotation using LabelImg

To train the XRAInet algorithm, we enlisted the assistance of two experienced pediatric surgeons who were blind to the clinical outcomes of the cases. The X-ray images were annotated using the LabelImg tool, an open-source graphical image annotation tool. The annotators, guided by their expertise in diagnosing and treating pneumothorax and pleural effusion, carefully drew bounding boxes around the regions of interest and labeled them accordingly.

To ensure the accuracy and consistency of the annotations, a third pediatric surgeon independently reviewed and verified each case. Any discrepancies or disagreements between the annotators were resolved through consensus agreement. This meticulous annotation process guarantees the reliability and validity of the annotated data, which is crucial for training the XRAInet algorithm effectively.

Data augmentation techniques, including random rotations, flips, and scaling, were applied to improve the generalization of XRAInet and prevent overfitting. The data set was split into training (80%) and validation (20%) sets, in accordance with original YOLO algorithm. The training set was used to train the XRAInet algorithm, while the validation set was utilized to evaluate its performance during training.

XRAInet was trained for 500 epochs using the Adam optimizer with a learning rate of 0.001. Mean Average Precision (mAP) served as the primary evaluation metric, with an IoU threshold of 0.5 to determine correct predictions. Training progress and model evaluation indicators, including precision, recall, and mAP@0.5, were monitored throughout the training process.

This study utilized multiple evaluation metrics to assess the performance of the model. These metrics included mAP, Recall, Precision, and F1 score. The choice of these metrics was driven by the nature of DL and the specific algorithm employed, YOLO (You Only Look Once). mAP provides an overall measure of recognition accuracy by calculating the average precision (AP) values across different categories. It takes into account both precision and recall, considering the precision at different recall levels. This allows for a comprehensive assessment of the model's performance across various categories. Precision measures the accuracy of positive category predictions. It calculates the ratio of correctly identified positive samples to the total number of samples predicted as positive. Recall, on the other hand, assesses the proportion of correctly identified positive samples out of the total number of actual positive samples. It indicates the model's ability to detect positive instances. The F1 score is the harmonic mean of precision and recall, providing a

10990496, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppul.27133 by Mustafa Alper Akay - Kocaeli Universitesi , Wiley Online Library on [04/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

balanced measure of the model's performance by considering both metrics simultaneously. Monitoring these metrics allows for informed decisions regarding the model's performance and identification of areas for improvement.

3 | RESULTS

We used 510 images from 170 patients for training and validation, including chest tube patients and conservative management group images. An additional real-time data set of 51 images patients requiring tube thoracostomy or conservative treatment was used for testing.

XRAInet's effectiveness was evaluated using a data set of X-ray images that required tube thoracostomy. The performance of XRAInet was assessed using several evaluation metrics, including precision–recall curve, precision-confidence curve, and F1 confidence curve, applied on an independent sample set. Additional metrics were also assessed to further evaluate the model.

The precision–recall curve analysis showed an overall mAP of 0.918 at an Intersection over Union (IoU) threshold of 0.5 (Figure 1). This indicates the algorithm's ability to accurately identify and localize regions indicative of pneumothorax and pleural effusion, requiring intervention. A higher mAP value signifies a stronger performance in distinguishing between positive and negative instances.

The diagnostic test demonstrated 64.00% sensitivity (95% confidence interval [CI]: 42.52%–82.03%) and 96.15% specificity (95% CI: 80.36%–99.90%). A positive result yielded a substantial positive likelihood ratio of 16.64 (95% CI: 2.38–116.29), while a negative result had a low negative likelihood ratio of 0.37 (95% CI: 0.22–0.63). With disease prevalence at 49.02%, the positive predictive value was 94.12% (95% CI: 69.60%–99.11%), and the negative predictive value was 73.53% (95% CI: 62.09%–82.49%). Overall accuracy was 80.39% (95% CI: 66.88%–90.18%). These results highlight the test's potential to identify positive cases effectively, along with its ability to accurately classify negative cases (Table 2). Sample diagnostic test results are provided in Figure 2.

4 | DISCUSSION

DL and AI have gained significant attention in the field of medical imaging, offering promising opportunities to improve diagnostic accuracy and patient care.[10] DL models excel at learning complex patterns and features from large datasets, enabling them to assist in various tasks such as detection, classification, and prediction in medical imaging applications.[11]

TABLE 2 Diagnostic performance of the model on an external test data set.

Statistic	Value	95% CI
Sensitivity	64.00%	42.52%–82.03%
Specificity	96.15%	80.36%–99.90%
Positive likelihood ratio	16.64	2.38–116.29
Negative likelihood ratio	0.37	0.22–0.63
Disease prevalence	49.02%	
Positive predictive value	94.12%	69.60%–99.11%
Negative predictive value	73.53%	62.09%–82.49%
Accuracy	80.39%	66.88%–90.18%

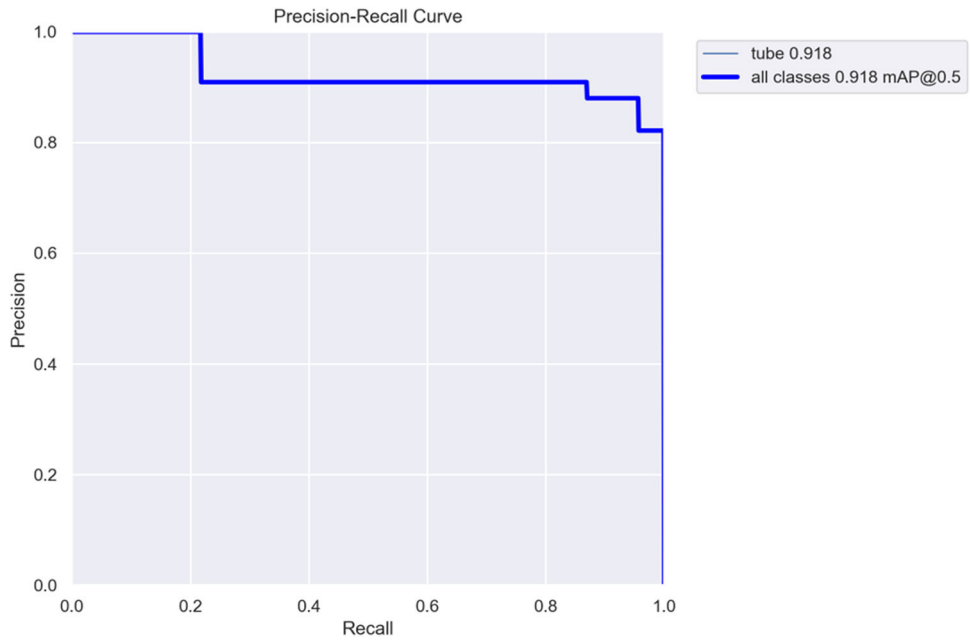


FIGURE 1 The precision–recall curve analysis. [Color figure can be viewed at wileyonlinelibrary.com]

10990496, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppul.27133 by Mustafa Alper Akay - Kocaeli Universitesi , Wiley Online Library on [04/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

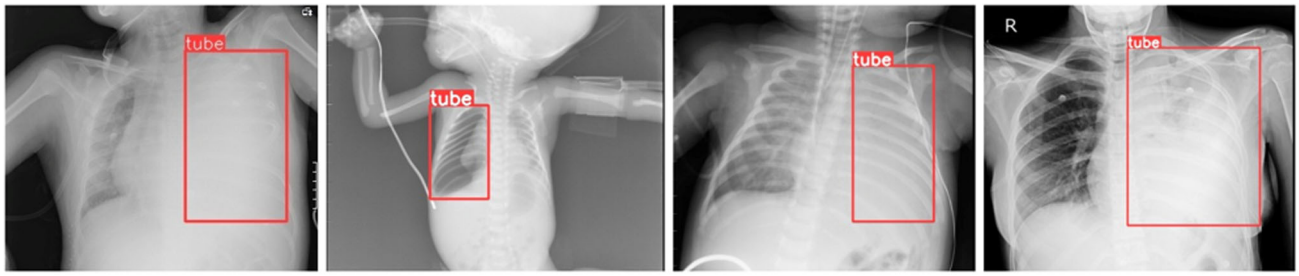


FIGURE 2 Sample diagnostic test results.

Numerous studies have explored the application of DL techniques for object detection in lung imaging, including the detection of pneumothorax and pleural effusion. For example, Taylor et al. tested a model for detecting pneumothorax.[12] They reported that their "high sensitivity model" achieved a sensitivity of 0.84, specificity of 0.90, and AUC of 0.94 in internal testing (95% CI: 0.78–0.90, 0.89–0.92, and 0.94, respectively). Their "high specificity model" showed a sensitivity of 0.80, specificity of 0.97, and AUC of 0.96 for the same testing (95% CI: 0.72–0.86, 0.96–0.98, and 0.96, respectively). PPVs were 0.45 and 0.71 for the high sensitivity and high specificity models, respectively (95% CI: 0.39–0.51 and 0.63–0.77). The models demonstrated decreased performance on the full set (sensitivity 0.55 and 0.45, specificity 0.90 and 0.97, and AUC 0.82 and 0.86, respectively). External testing using the NIH data set showed further performance decline (sensitivity 0.28–0.49, specificity 0.85–0.97, and AUC 0.75). In another study by Hillis et al., their AI model demonstrated high accuracy in detecting pneumothorax and tension pneumothorax.[13] The model achieved an area under the receiver operating characteristic curve (AUC) of 0.979 (95% CI: 0.970–0.987) for detecting any pneumothorax, with a sensitivity of 94.3% (95% CI: 92.0%–96.3%) and specificity of 92.0% (95% CI: 89.6%–94.2%). Furthermore, for tension pneumothorax, the model achieved an AUC of 0.987 (95% CI: 0.980–0.992), with a sensitivity of 94.5% (95% CI: 90.6%–97.7%) and specificity of 95.3% (95% CI: 93.9%–96.6%). These studies demonstrate the effectiveness of DL in lung imaging applications.

In our study, we aimed to investigate the effectiveness of the XRAInet algorithm in detecting and providing decision support for both pneumothorax and pleural effusion cases in X-ray images. Unlike existing literature, our research encompassed the identification of these conditions as well as the determination of surgical intervention requirements. The results of our study revealed that XRAInet achieved a mAP of 0.918 at an IoU threshold of 0.5. This demonstrates its effectiveness in accurately identifying and localizing regions indicative of pneumothorax and pleural effusion, while also providing valuable decision support regarding the need for surgical intervention. In the external testing data set consisting of 25 tubes requiring and 26 conservative management requiring patiens, our model demonstrated 64.00% sensitivity (95% CI: 42.52%–82.03%) and 96.15% specificity (95% CI: 80.36%–99.90%) regarding choosing the right surgical decision. Our model, while undertaking a more complex task, still demonstrates competitive performance, making its sensitivity acceptable in the context of our study's goals. We acknowledge the absence of similar studies for a direct comparison, but when juxtaposed with studies addressing less complex tasks, our model holds up well, reinforcing the acceptability of its sensitivity level.

It is worth noting that the choice of imaging modality plays a crucial role in the performance of DL models. CT scans provide higher-resolution images and are considered the gold standard for evaluating lung pathologies.[9] However, X-ray imaging is more widely accessible, cost-effective, and involves reduced radiation exposure compared to CT scans. Our study focused on X-ray imaging due to its practicality and widespread use in clinical settings. Nevertheless, future research could explore the integration of hybrid models that combine both X-ray and CT imaging to leverage the strengths of each modality.

The deployment of DL models in clinical settings requires addressing several considerations. High-quality annotated data sets are essential for training DL models effectively. The annotation process should involve experts, both surgeons and radiologists, with expertise in diagnosing and treating pneumothorax and pleural effusion. Additionally, ensuring the accuracy and consistency of annotations is crucial. A second radiologist review and verification process can help validate the annotations and enhance the reliability of the data set.

Interpretability and transparency are also important factors when deploying DL models in clinical practice. DL models are often considered as black boxes, making it difficult to understand their decision-making process. Developing explainable DL models is a growing area of research, aiming to provide insights into the reasoning behind the predictions, enhance trust, and facilitate the integration of DL models into clinical workflows.

Moreover, regulatory considerations are essential to ensure the safe and effective use of DL models in medical imaging. Regulatory bodies, such as the FDA, are actively developing guidelines and frameworks to evaluate and regulate DL-based medical devices. Compliance with these regulations is crucial to guarantee patient safety, data privacy, and ethical considerations in deploying DL models in healthcare settings.[14]

10990496, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppul.27133 by Mustafa Alper Akay - Kocaeli Üniversitesi , Wiley Online Library on [04/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 | CONCLUSION

In conclusion, our study demonstrates the potential of the XRAInet algorithm in detecting and providing decision support for pneumothorax and pleural effusion cases in X-ray images. The results align with previous research highlighting the effectiveness of DL models in lung imaging. However, further validation, refinement, and evaluation are necessary before the integration of the XRAInet algorithm into routine clinical practice. Future research directions may include expanding the data set, exploring hybrid models that combine multiple imaging modalities, addressing data quality challenges, improving interpretability, and adhering to regulatory requirements.

AUTHOR CONTRIBUTIONS

Mustafa Alper Akay: Conceptualization; methodology; data curation; project administration; resources; writing—review and editing; writing—original draft; investigation; supervision. **Ozan Can Tatar**: Conceptualization; methodology; software; investigation; validation; formal analysis; resources; visualization; writing—original draft. **Elif Tatar**: Data curation; software; investigation; validation; writing—original draft. **Beyza Nur Tağman**: Data curation; investigation. **Semih Metin**: Data curation; investigation. **Onursal Varlıklı**: Data curation.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Mustafa Alper Akay http://orcid.org/0000-0003-3315-6098
Ozan Can Tatar http://orcid.org/0000-0002-9046-7362
Elif Tatar http://orcid.org/0000-0003-0337-9362
Beyza Nur Tağman http://orcid.org/0009-0001-1068-0877
Semih Metin http://orcid.org/0000-0001-6356-6288
Onursal Varlıklı http://orcid.org/0000-0001-8714-1874

REFERENCES

1. Harris M, Rocker J. Pneumothorax in pediatric patients: management strategies to improve patient outcomes. *Pediatr Emerg Med Pract*. 2017;14(3):1-28.
2. Cashen K, Petersen TL. Pleural effusions and pneumothoraces. *Pediatr Rev*. 2017;38(4):170-181. doi:10.1542/pir.2016-0088
3. Irmici G, Cè M, Caloro E, et al. Chest x-ray in emergency radiology: what artificial intelligence applications are available. *Diagnostics*. 2023;13(2):216. doi:10.3390/diagnostics13020216
4. Schalekamp S, Klein WM, van Leeuwen KG. Current and emerging artificial intelligence applications in chest imaging: a pediatric perspective. *Pediatr Radiol*. 2022;52(11):2120-2130. doi:10.1007/s00247-021-05146-0
5. Ahn JS, Ebrahimian S, McDermott S, et al. Association of artificial intelligence-aided chest radiograph interpretation with reader performance and efficiency. *JAMA Netw Open*. 2022;5(8):e2229289. doi:10.1001/jamanetworkopen.2022.29289
6. Rudolph J, Huemmer C, Ghesu FC, et al. Artificial intelligence in chest radiography reporting accuracy: added clinical value in the emergency unit setting without 24/7 radiology coverage. *Invest Radiol*. 2022;57(2):90-98. doi:10.1097/RLI.0000000000000813
7. Anderson D, Chen SA, Godoy LA, Brown LM, Cooke DT. Comprehensive review of chest tube management: a review. *JAMA Surg*. 2022;157(3):269-274. doi:10.1001/jamasurg.2021.7050
8. Terven J, Cordova-Esparza D. A Comprehensive Review of YOLO: From YOLOv1 and Beyond. Published online June 9 2023. doi:10.48550/arXiv.2304.00501
9. Tækker M, Kristjánsdóttir B, Andersen MB, et al. Diagnostic accuracy of ultra-low-dose chest computed tomography in an emergency department. *Acta Radiol*. 2022;63(3):336-344. doi:10.1177/0284185121995804
10. Chen X, Wang X, Zhang K, et al. Recent advances and clinical applications of deep learning in medical image analysis. *Med Image Anal*. 2022;79:102444. doi:10.1016/j.media.2022.102444
11. Duggento A, Conti A, Mauriello A, Guerrisi M, Toschi N. Deep computational pathology in breast cancer. *Sem Cancer Biol*. 2021;72:226-237. doi:10.1016/j.semcancer.2020.08.006
12. Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: a retrospective study. *PLoS Med*. 2018;15(11):e1002697. doi:10.1371/journal.pmed.1002697
13. Hillis JM, Bizzo BC, Mercaldo S, et al. Evaluation of an artificial intelligence model for detection of pneumothorax and tension pneumothorax in chest radiographs. *JAMA Netw Open*. 2022;5(12):e2247172. doi:10.1001/jamanetworkopen.2022.47172
14. Potnis KC, Ross JS, Aneja S, Gross CP, Richman IB. Artificial intelligence in breast cancer screening: evaluation of FDA device regulation and future recommendations. *JAMA Intern Med*. 2022;182(12):1306-1312. doi:10.1001/jamainternmed.2022.4969

How to cite this article: Akay MA, Tatar OC, Tatar E, Tağman BN, Metin S, Varlıklı O. XRAInet: aI-based decision support for pneumothorax and pleural effusion management. *Pediatr Pulmonol*. 2024;1-6. doi:10.1002/ppul.27133

10990496, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ppul.27133 by Mustafa Alper Akay - Kocaeli Üniversitesi , Wiley Online Library on [04/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License