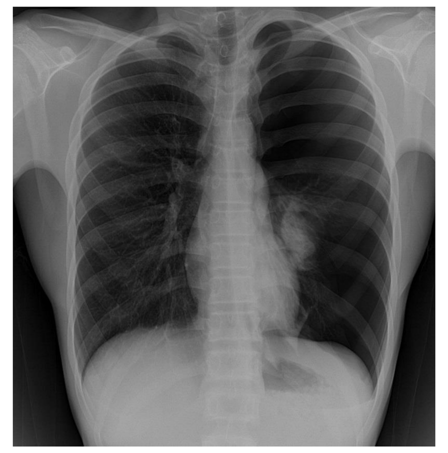
In-Depth Comparative Analysis of Commercial Pneumothorax Detection Tools on Plain Chest Radiographs: A Multivendor Diagnostic Accuracy Study

Dr Katrina Nash[1,2,3], Dr James Vaz[1,3], Dr Abdala Espinosa[3], Professor Alex Novak[1,3], Dr Sarim Ather[1,3,4]

1-John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust; 2-Oxford University Clinical Academic Graduate School, University of Oxford, Oxford; 3 – Oxford Clinical Artificial Intelligence Research (OxCAIR), Oxford University Hospitals, Oxford; 4-RAQIC Ltd, Oxford; 5 - Emergency Medicine Research Oxford (EMROx), Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Background

- CXR are the most common radiograph[1]
- Significant backlog of CXR reports due to radiologist shortages[2]
- Clinicians are often required to interpret CXR to make initial management and discharge decisions[3]
- CXR play a key role in the diagnosis of pneumothorax, a potentially fatal diagnosis[3]
- Artificial intelligence has emerged as a promising tool to aid in CXR interpretation and reduce reporting backlogs[2,4]

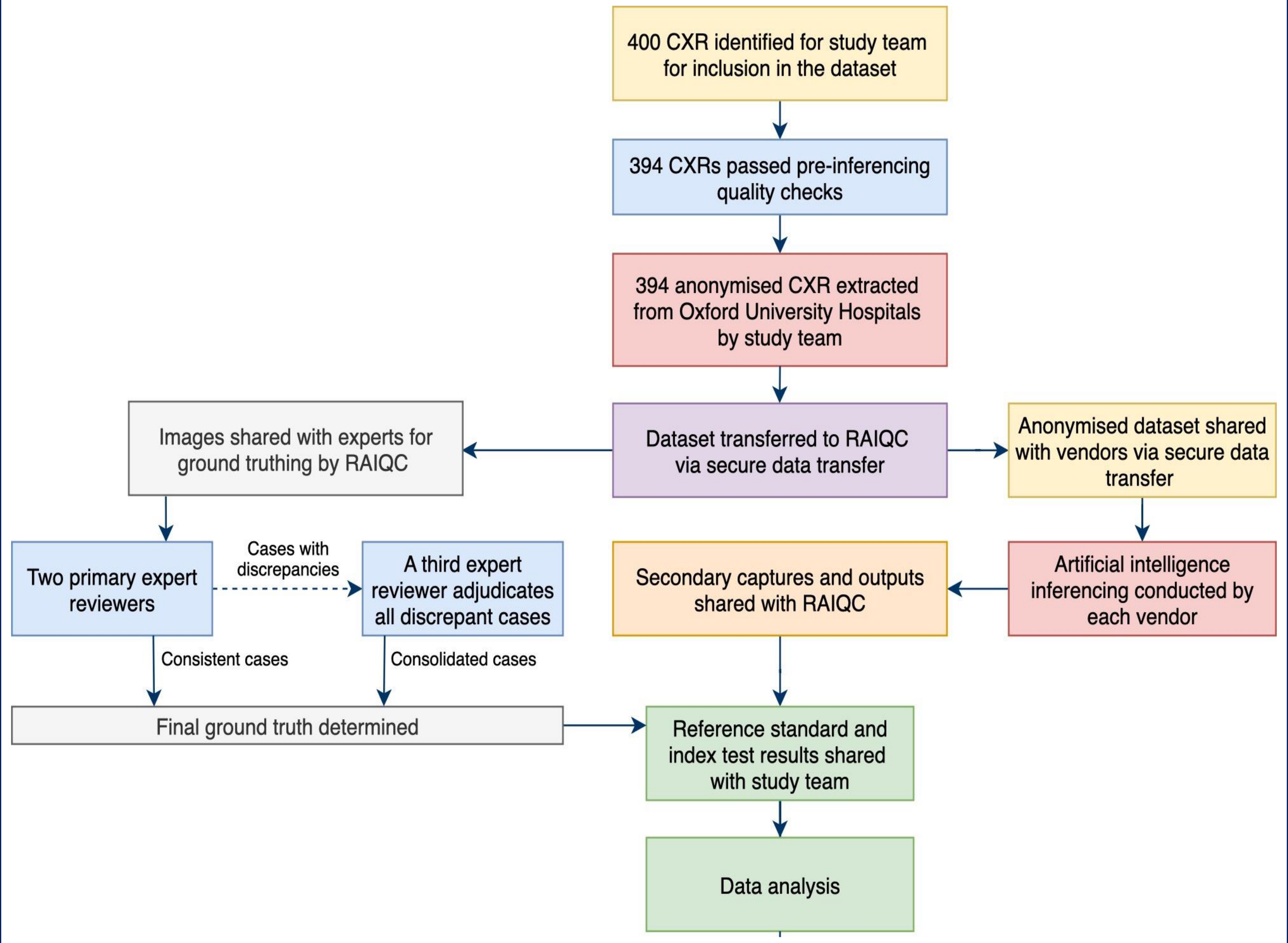


Case courtesy of David Puyó Vera, Radiopaedia.org, rID: 53786

Aim: To assess the accuracy of multiple AI devices in the detection of pneumothorax on CXR, and evaluate the impact of demographic, clinical and imaging factors

Methodology

- Retrospective evaluation of nine CE marked AI devices for the detection of pneumothorax on CXR
- All CE-marked devices as of October 2024 were invited to partake



Results

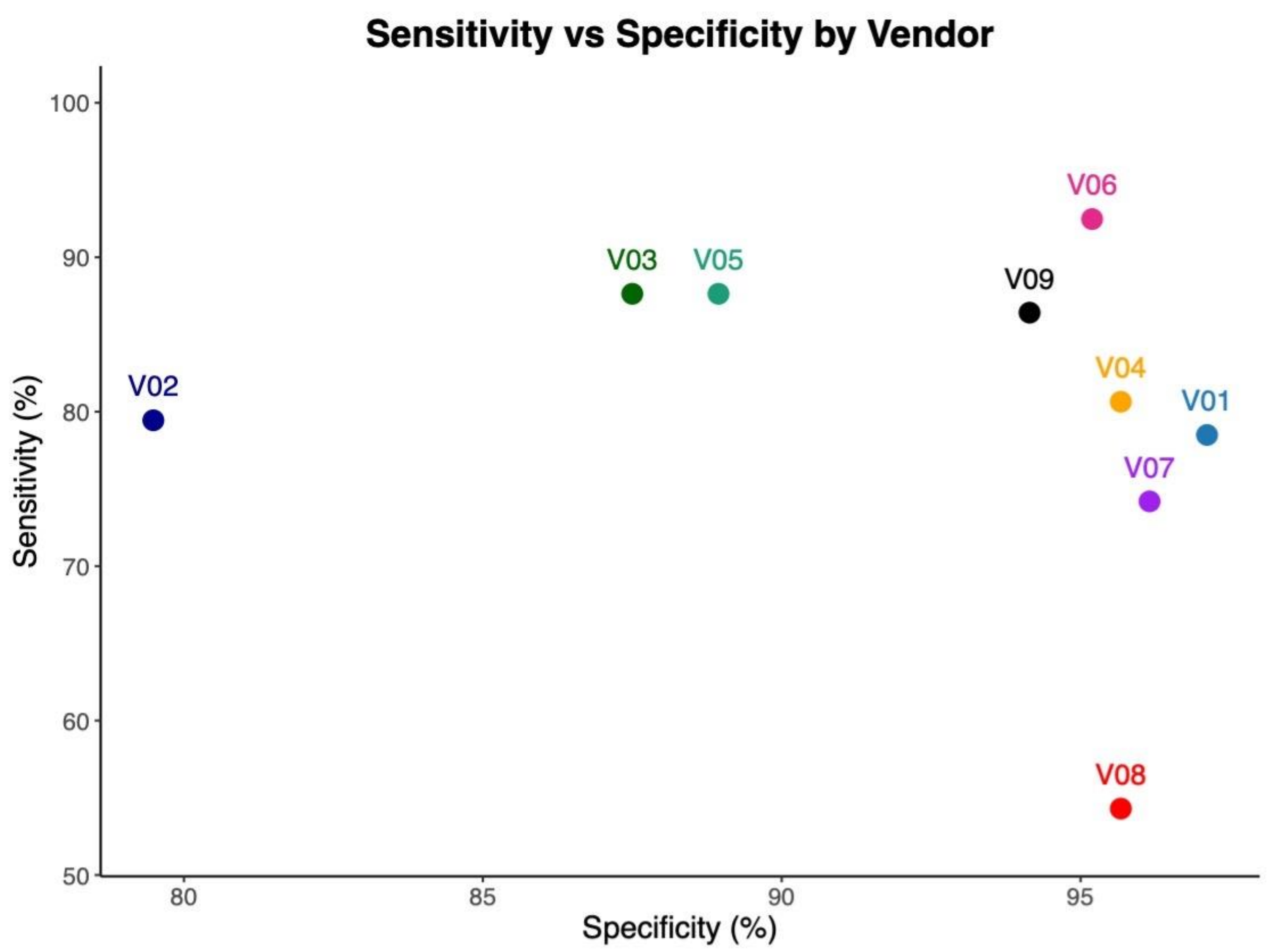


Figure 1: Scatter graph showing sensitivity vs. specificity by vendor

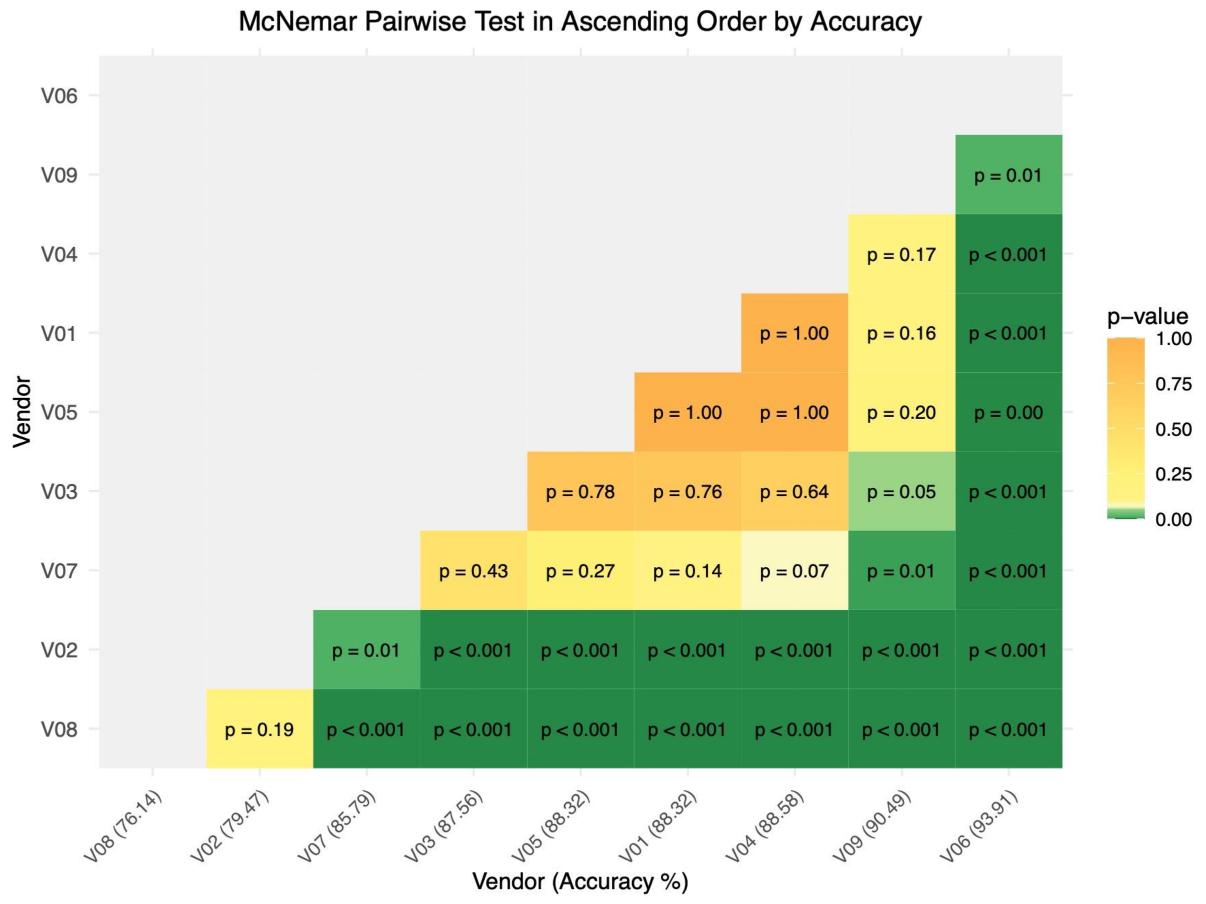


Figure 2: Heatmap with p-value for pairwise comparisons between vendors for a statistically significant difference in performance

Demographics

- Mean age: 56 years old (SD 21)
- 156 females (39.6%); 238 males (60.4%)
- 186 positive(47.2%) and 208 (52.8%) negative for pneumothorax
- 135 (35%) AP and 256 (65%) PA radiographs
- 324 (82.2%) fixed CXR and 69 *17.5%) mobile CXR

V01 = OXIPIT
V03 = VUNO
V04 = Lunit
V05 = RAYSCAPE
V06 = harrison.ai
V09 = Gleamer
V02, V07 and V08 = anonymous

*7 vendors were able to successfully inference all cases. V02 failed to inference 19 cases and V09 failed to inference 5 cases

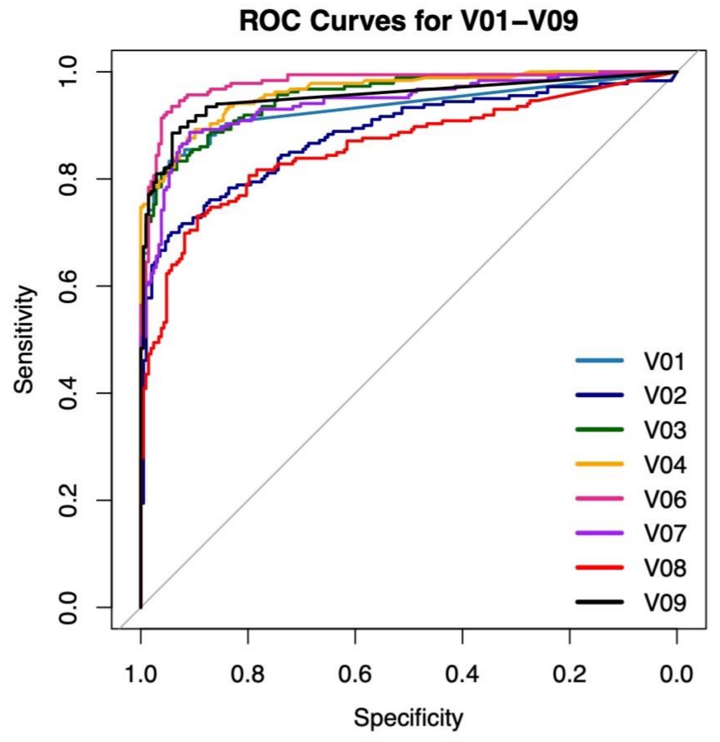


Figure 3: Receiver Operating Characteristic (ROC) Curve by vendor

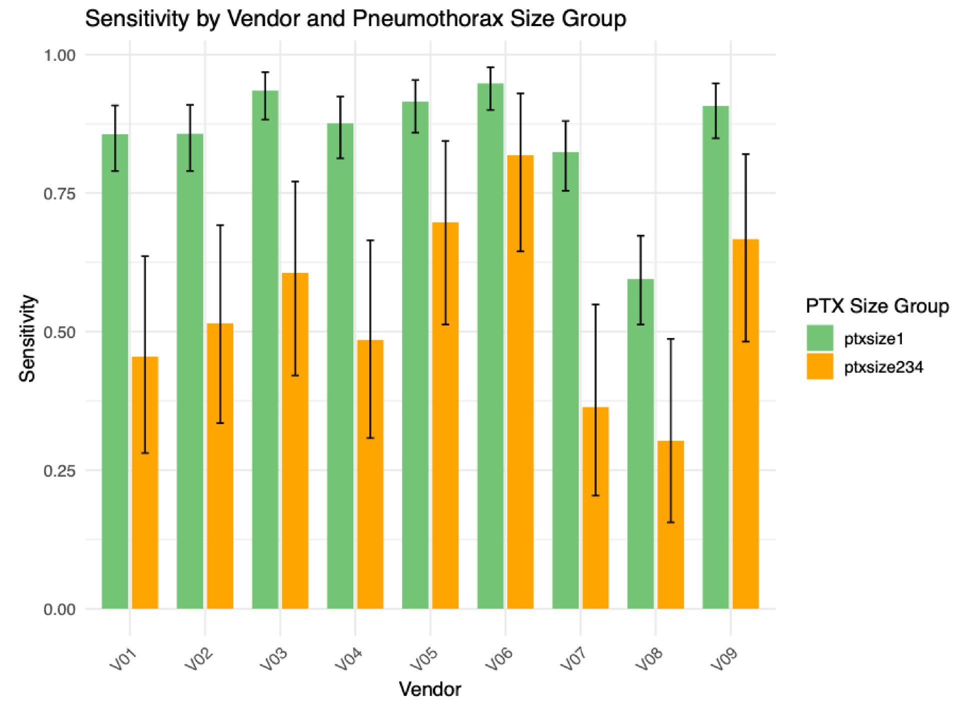


Figure 4: Sensitivity by pneumothorax size for each vendor (ptxsize 1 = <10mm, ptxsize2,3,4 >10mm)

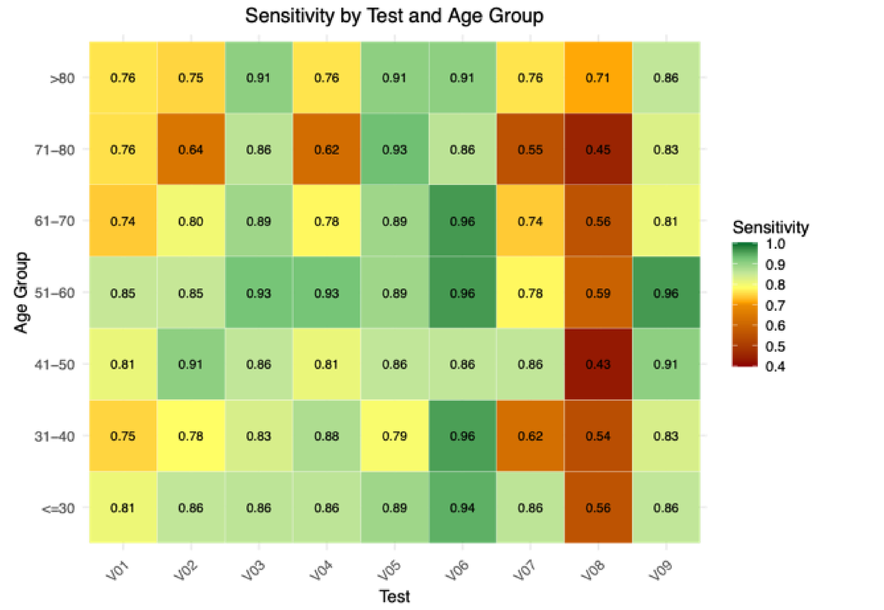


Figure 5: Heatmap demonstrating sensitivity by age group for each vendor

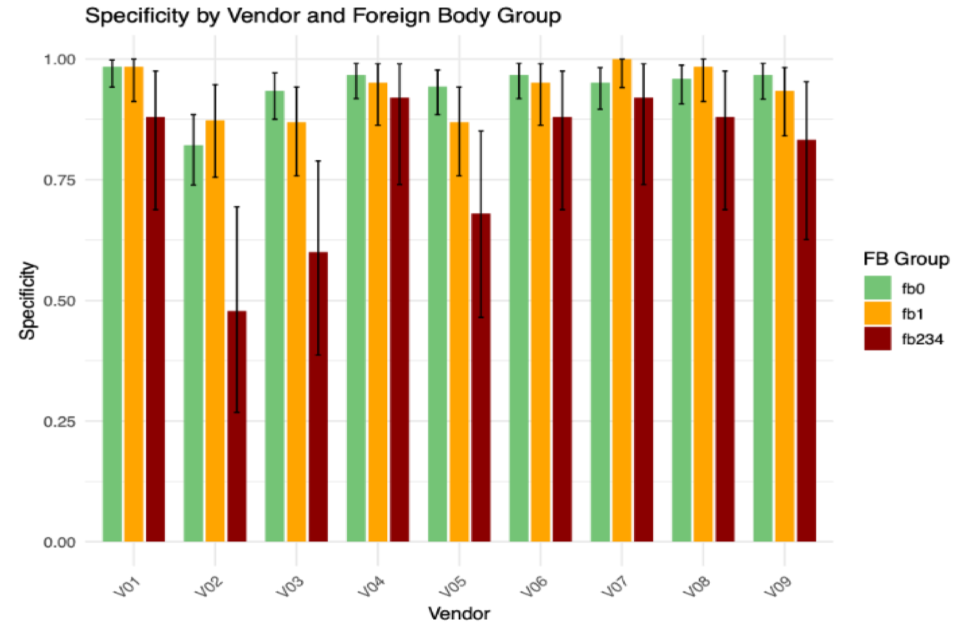


Figure 6: Specificity by number of foreign bodies present on CXR by vendor

Strengths and Limitations

Strengths

- ✓ Largest multi-vendor AI comparison study
- ✓ All CE approved AI vendors
- ✓ Robust ground truth definition
- ✓ Enriched dataset
- ✓ Established methodology

Limitations

- Small datasets → further larger studies should be studied
- "Out of the box" → local fine tuning may improve performance
- Single pathology → results may not translate to other pathologies

Conclusions

High accuracy of AI devices in the detection of pneumothorax

- Specificity exceeded sensitivity → higher risk of false negatives
- Significant variation across the group, with clusters of similarly performing devices → AI devices are not created equally

Patterns of limitations seen in AI devices

- Reduced sensitivity for smaller pneumothoraces → more likely to be missed
- Specificity for CXRs with concurrent foreign bodies → more likely to be misdiagnosed by AI
- Differences in performance by age group → informs clinical deployment & highlights opportunity for further training

Future prospective multi-site evaluations should be undertaken

- Explore AI device limiations to inform clinical safety and algorithm training
- Understand operational impact
- Evaluate the peformance of AI devices within the intended population

References

1. Jones CM, Buchlak QD, Oakden-Rayner L, et al. Chest radiographs and machine learning: past, present and future. J Med Imaging Radiat Oncol. 2021;65:538.
2. The Royal College of Radiologists. Clinical radiology workforce census 2023 [Internet]. London: The Royal College of Radiologists; 2024. Available from: https://www.rcr.ac.uk/media/5befglss/rcr-census-clinical-radiology-workforce-census-2023.pdf
3. Zarogoulidis P, Kioumis I, Pitsiou G, et al. Pneumothorax: from definition to diagnosis and treatment. J Thorac Dis. 2014;6:S372–6.
4. Katzman BD, Alabousi M, Islam N, Zha N, Patlas MN. Deep learning for pneumothorax detection on chest radiograph: a diagnostic test accuracy systematic review and meta-analysis. Can Assoc Radiol J. 2024;75:525–33.