

Can artificial intelligence algorithm reliably distinguish normal chest X-rays from abnormal chest X-rays

A real-world post marketing surveillance analysis

Authors: Dr Neelan Das, Dr Mohammed Elbargesy, Steve Morris, Dennis Robert, Dhruv Shah, Banu Bilgili, Ansh Vazirani

qure.ai

NHS East Kent Hospitals University NHS Foundation Trust

Purpose

Only about 23-26% of the chest X-ray (CXR) examinations conducted are reported on the same day in the NHS.[1] Artificial intelligence (AI) driven chest X-ray (CXR) segregation into normal and abnormal CXR has the potential to optimize radiologists' workflow and productivity.[2,3] We investigated the accuracy of a regulatory cleared AI software device (qXR) in a real-world setting as part of a post marketing surveillance activity.

Results

After excluding duplicates and those with invalid CXR reports, 8313 CXRs from A&E patients (6273 [75.5%] > 50 years of age; 4184 [50.3%] females) were analysed. AI classified 2637 (31.7%) CXRs as normal of which 243 were FN CXRs.

As per the radiologist panel consensus, 126 (51.9%) of the FN CXRs were normal, 78 (32.1%) had inconsequential findings, 37 (15.2%) had consequential findings and 2 (0.08%) contained critical findings. This indicated an NPV of 95.6% (95% CI: 94.7-96.3) of AI when classifying normal CXRs. The sensitivity and specificity of AI in detecting abnormal CXRs was 96.7% (96.0-97.2) and 52.5% (51.1-54.0), respectively.

After including only consequential and critical findings as significant FN CXRs, the NPV was found to be 98.5% (98.0-98.9).

The two CXRs with critical missed findings were reported to have hilar enlargement and consolidation in the presence of hardware as per the panel.

Methods and materials

Consecutive frontal CXRs acquired from A&E patients aged > 18 years during Nov-2023 to Feb-2024 were used for this analysis. CXRs with no detected findings were classified as normal CXRs by AI. The original radiologist report (ORR) of CXRs was also used to classify a CXR as abnormal or normal depending on whether ORR reported an abnormal finding. Abnormal CXRs as per ORR, but normal by AI were treated as false negative (FN) CXRs. FN CXRs were reviewed by three independent blinded radiologists to categorize the CXRs into critical (immediate attention required), consequential (follow-up within 48 hours required), and inconsequential (no expedited follow-up required), based on the criticality of CXR findings as per consensus. Negative predictive (NPV), sensitivity and specificity of AI in detecting abnormal CXRs were calculated.

Conclusion

AI was able to confidently classify normal CXRs with high NPV. Such a tool may have the potential to optimize the radiologists' workflow and productivity. Suboptimal specificity resulting in a significant proportion of false positive findings is an area of improvement for AI.

References

1. NHS England. Diagnostic imaging dataset—statistical release [Internet]. Version 1. 2024. Accessed Dec 9, 2024. https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2024/11/Statistical-Release-21st-November-2024-2.pdf

2. Harolds JA, Parikh JR, Bluth EI, Dutton SC, Recht MP. Burnout of Radiologists: Frequency, Risk Factors, and Remedies: A Report of the ACR Commission on Human Resources. J Am Coll Radiol. 2016 Apr;13(4):411-6. doi: 10.1016/j.jacr.2015.11.003. Epub 2016 Jan 4. PMID: 26768546.

3. Kumar A, Patel P, Robert D, Kumar S, Khetani A, Reddy B, Srivastava A. Accuracy of an artificial intelligence-enabled diagnostic assistance device in recognizing normal chest radiographs: a service evaluation. BJR Open. 2024 Sep 14;6(1):tzae029. doi: 10.1093/bjro/tzae029. PMID: 39350939; PMCID: PMC11441651.