AI as a Medical Device for autonomous reporting of chest radiographs: a large-centre retrospective validation.

Dr. Aditya U Kale
Dr. Sonam Vadera
Mr. Robert McEwan
Dr. Zaira Akhtar
Mr. Barnaby Waters
Dr. Akif Malik
Dr. Ben Holloway
Dr. Deborah Tattersall
Prof. Alastair Denniston
Dr. Gareth Lewis
Assoc Prof. Xiaoxuan Liu

University Hospitals Birmingham NHS Foundation Trust, UK

National Institute for Health and Care Research Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK

Background

Chest radiography is a commonly used imaging modality, playing an important role in the diagnosis of a wide range of medical conditions. The ease of access to x-ray scanners in secondary care settings means that chest-rays are some of the most frequently requested scans. Due to the increasing volume of radiological imaging in the context of limited resource, there has been interest in Artificial Intelligence (AI) enabled image classification.

ChestLink (Oxipit UAB, Vilinius, Lithuania) is a class IIa CE marked AI as a Medical Device (AIaMD), intended to reduce radiology workload through autonomous reporting of normal chest X-rays. ChestLink is able to determine which scans are normal, and subsequently generate a report for these scans. ChestLink has been validated through two retrospective analyses demonstrating an autonomous reporting rate of 28% and sensitivity of over 99%. As with all AIaMD there will be generalisability limitations to be considered. ChestLink has not been deployed in an NHS setting, however there is interest in its use for autonomous reporting for gains in allocative efficiency.

University Hospitals Birmingham NHS Foundation Trust (UHB) is currently exploring the procurement of ChestLink. The need for recurring local validation has been proposed as an alternative to external validation of AIaMD, aligning with post-market surveillance requirements for medical devices. The Medical Algorithmic Audit is a safety monitoring framework which enables safety monitoring through a collaborative process involving multiple stakeholder groups. This framework is being used by UHB as part of the implementation process to identify potential failure modes and plan appropriate mitigations. The audit will inform the next steps for the implementation process including planning of a prospective silent trial.

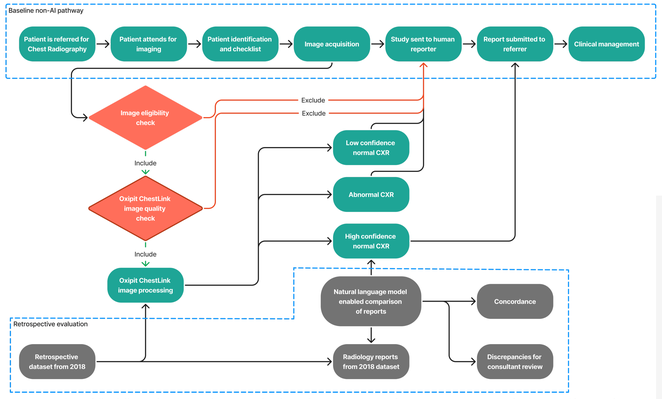


Figure 1: Simplified flow diagram showing the baseline non-AI pathway, new steps with Oxipit ChestLink implementation, and an overview of the local retrospective validation.

Aim: to evaluate the local performance of ChestLink, a CE-marked Class IIb AI as a Medical Device developed by Oxipit (Vilnius, Lithuania), designed to autonomously report high confidence normal chest X-ray studies.

Methods

A retrospective analysis was undertaken at University Hospitals Birmingham NHS Foundation Trust, examining 197,258 chest X-rays across GP, A&E, outpatient, and inpatient referral sources. A separately validated natural language processing (NLP) tool is provided by Oxipit to classify the clinician reports. The concordance between ChestLink's reports and the clinician reports was assessed. Discordant cases were reviewed by a radiology consultant who classified any missed findings. Images were classified using a schema: 1) critical finding, 2) non-critical but urgent, or 3) clinically significant (but neither critical nor urgent) and 4) unremarkable.

Scoping	Mapping	Artefact Collection	Testing	Reflection	Post audit
Audit scope: All CXRs obtain over 2018 calendar year	**AI system:** Deep Learning based AI system	**Audit checklist:** • Intended use statement • Intended impact statement • FMEA clinical task risk analysis • FMEA risk priority number document • Datasets • Data description • Data, including explainability artefacts • Data flow diagram • AI model itself • Model summary • Previous evaluation materials	**Overall Performance:** 23.4% of normal scans reported with over 99% sensitivity	*Recommendations TBC	**Algorithmic audit summary:** Report currently in draft for dissemination to all key stakeholders within hospital
Intended use: Autonomous generation of CXR reports for high confidence normals	**Health-care task:** Analysis and reporting of Chest Radiographs		**Exploratory error analysis:** 593 discrepant cases audited	**Developer actions:** Modifications to normal definitions as required	
Intended impact: Radiology workload reduction	**Personnel and resources**		**Subgroup testing:** Age, sex, ethnicity and referral source	**Deployer actions:** Audit planning and accommodating training requirements Selection of referral sources to maximise value	**Plan re-audit:** Medical Algorithmic Audit framework is being built into local governance processes
	Identify and prioritise risks: Risk Priority Numbers		**Adversarial testing:** no formal testing conducted as yet		
Failure Modes and Effects Analysis (FMEA)					

The Medical Algorithmic Audit (MAA) framework was used to structure this analysis. The MAA is a safety monitoring framework for algorithmic error auditing and failure mode detection. The framework involves a six stage SMACTR approach starting with "Scoping" and ending with "Reflection" and post audit actions.

Figure 2 (Left): Overview of MAA as applied to Oxipit ChestLink validation. A dataset of CXRs was curated for the calendar year 2018. We conducted a failure modes and effects analysis to identify high priority risks. This enabled development of the test protocol, in addition to the implementation plan. Demographic summaries will be created to assess for differential reporting success across subgroups. Finally, this data will be used to plan appropriate adversarial testing strategies.

Results

141,597 chest x-rays were within scope for ChestLink processing, of which 63,442 were normal scans based on the NLP analysis. ChestLink was able to confidently report 14,820 (23.4%) of these as normal, effectively removing them from further assessment and expediting the workflow. The highest autonomous reporting rate was observed in the GP setting, with ChestLink reporting 29.5% of normal studies. Conversely, autonomous reporting rate was lowest among inpatients, where it reported 15.1% of normal chest X-rays. Overall, discordance was observed in 593 (4.0%) cases, whereby ChestLink classified the X-ray as normal while radiologists reported abnormalities. Among these discordant cases 141 (0.075%) were labelled as abnormal during consultant review, of which 3 radiographs contained critical findings and 29 contained non-critical but urgent findings. 75 cases were clinically significant although not critical or urgent, and 34 cases had clinically unremarkable findings. Critical findings included two cases of pneumoperitoneum, and one pneumomediastinum. Urgent findings all included missed lung masses/nodules.

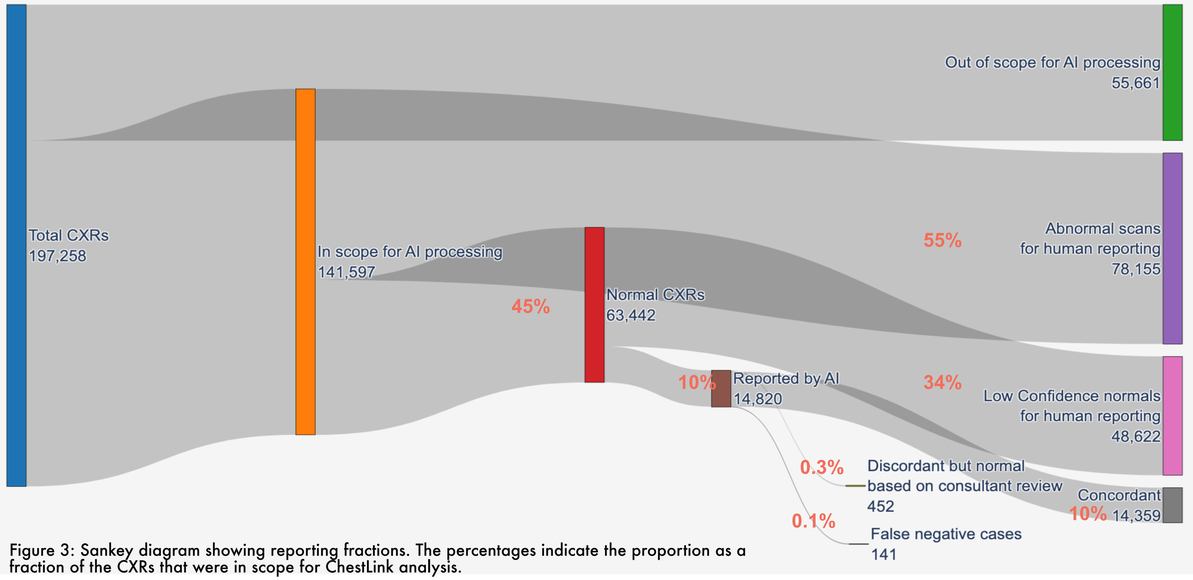


Figure 3: Sankey diagram showing reporting fractions. The percentages indicate the proportion as a fraction of the CXRs that were in scope for ChestLink analysis.

Discussion

ChestLink demonstrates promising accuracy in identifying normal chest X-rays, potentially reducing reporting workload of normal scans by nearly a quarter. The findings highlight the potential of AI tools like ChestLink in enhancing the efficiency of chest X-ray reporting. Integrating such AI tools into clinical workflows can aid radiologists in prioritizing abnormal cases, thereby facilitating timely patient care, and enabling re-allocation of clinician time. This retrospective analysis provides a foundation to plan for the next stage of implementation which will be a prospective silent implementation trial. Further testing will also allow for planning of the clinical pathway including which referral sources will be prioritised for ChestLink analysis.

@A_U_Kale
e. a.kale@bham.ac.uk