Real-World Multi-Country Evaluation of AI-Assisted vs Unaided Radiologist Performance in Chest X-Ray Reporting Using Rology's XGen AI Tool

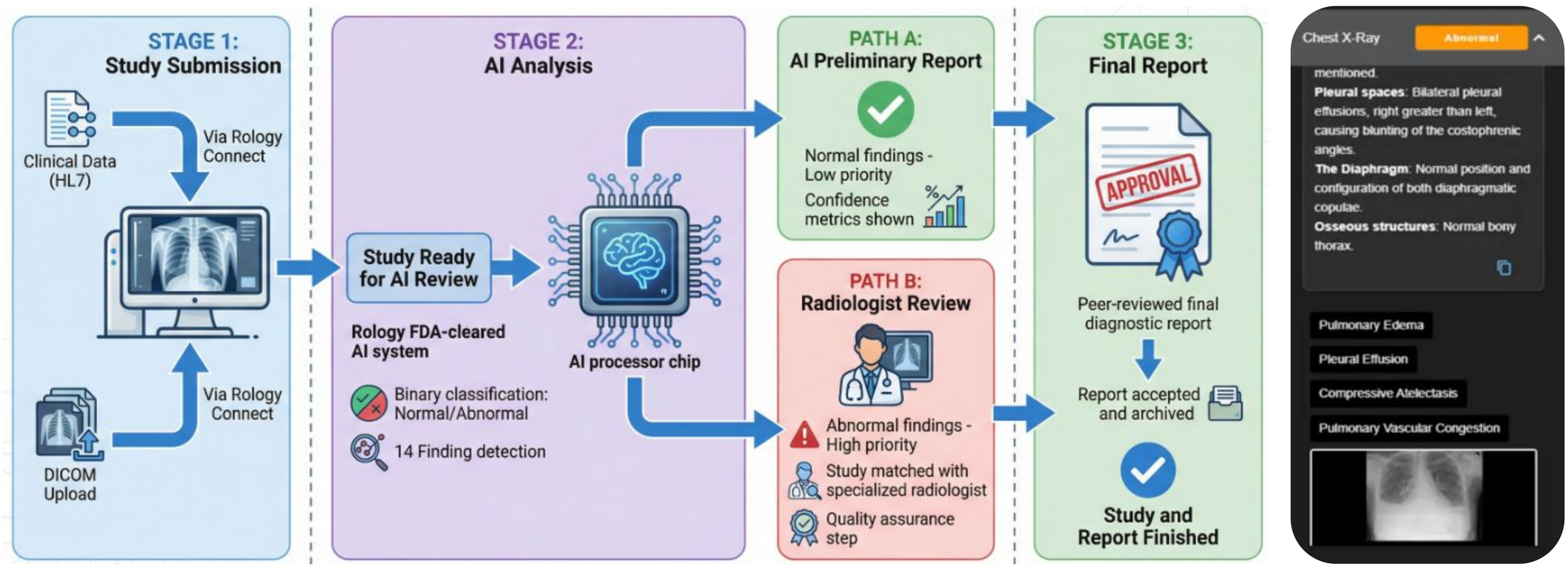
Sameh Ahmed[1], Ibrahim El-Shehaby[2], Naglaa Omar[3]

Introduction and Aim

Chest radiography is the most widely used imaging modality globally, yet variability in interpretation and limited radiology capacity persist, especially in emerging markets. This study evaluates the impact of an AI-powered chest X-ray (CXR) grounded reporting tool (XGen), a multimodal agentic AI system designed for disease detection, identification, and reporting across six key anatomical sections (osseous, diaphragm, pleura, heart, mediastinum, and lung), by comparing unaided radiologist interpretation with AI-assisted performance in real-world clinical settings.

Methods

A retrospective analysis was conducted on 2,824 chest X-ray studies from healthcare facilities across Egypt, Kenya, and Saudi Arabia, with 30% normal and 70% abnormal cases. The original reports generated on the Rology platform by a certified radiologist served as the ground truth. Two independent radiologists reviewed all cases, extracting abnormality labels, evaluating AI-predicted findings, and assessing both organ-level accuracy and full AI-generated reports. Performance was measured using sensitivity, specificity, precision (PPV), F1 score, and inter-rater agreement (Cohen's kappa).

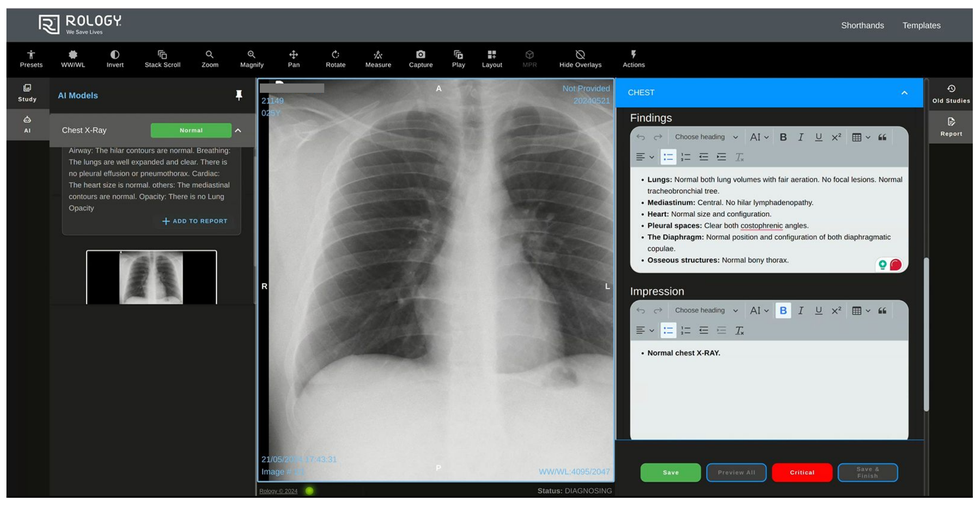


Results

AI achieved high specificity (96.7%) and moderate precision (53.7%), indicating strong performance in correctly identifying normal cases and improving reliability of positive predictions. Sensitivity was 50.6%, highlighting opportunities for improvement in detecting abnormalities. The global acceptance rate of AI-generated suggestions by radiologists was 87.3%, demonstrating strong clinical usability. Inter-radiologist agreement was moderate (κ = 0.536). Performance varied across anatomical regions, with highest agreement in osseous findings (F1 = 0.75) and lower performance in complex lung pathologies such as lesions and edema. Category-level analysis showed stronger AI alignment in common findings like cardiomegaly and pleural effusion.

Conclusion

AI-assisted reporting demonstrates strong potential to enhance diagnostic consistency, improve specificity, and support radiologists in real-world workflows. While sensitivity and performance on complex findings require further optimization, the high acceptance rate underscores clinical trust. These results support the integration of AI as an assistive tool to augment radiology capacity and standardize reporting, particularly in resource-constrained settings.



[1] *Mathematics Department, Faculty of Science, Cairo University, Egypt* sameh@sci.cu.edu.eg [2,3] *Rology*