An AI Assurance and Audit Workbench for Chest X-ray Models

Nattakrit Tadrabiab[1], Pasavee Jaipain[1] , Atiphon kaeochamnong[1*]

Thapanawat Chooklin[1]

[1]Princess Chulabhorn Science High School Nakhon si thammarat , Phanang, Bang Chang, Mueang Nakhon si thammarat, 80330 Thailand

**E-mail:6405723@pccnst.ac.th , 6706408@pccnst.ac.th , 6405702@pccnst.ac.th*

Abstract

CeXaR is a working, tested audit pipeline for chest X-ray artificial intelligence (AI) models, developed to address whether a diagnostic AI model that performs well in one hospital setting continues to perform reliably in another, given differences in imaging equipment, patient populations, and acquisition protocols. Rather than functioning as a diagnostic tool itself, CeXaR operates as an assurance and audit workbench that evaluates whether an externally developed model has been validated for a specific deployment context. The system implements a complete pipeline consisting of DICOM ingestion, de-identification, model inference, statistical auditing, and automated Evidence Report generation, and was validated end-to-end using real hospital-style DICOM data from a public chest X-ray dataset (VinDr-CXR, via Kaggle). Initial testing on 26 real files revealed a complete ingestion failure, later diagnosed as a DICOM-standard conformance gap; following remediation, the corrected pipeline achieved full acceptance on the same files and was subsequently scaled to a 53-file run. Reproducibility was demonstrated live: two independent runs on identical input data produced byte-for-byte identical result files, verified through content-derived cryptographic fingerprinting. A total of 109 automated tests across 18 source modules confirm system reliability. Because the eligibility criteria for statistically valid accuracy reporting were not met at this sample size, CeXaR correctly withheld point-estimate accuracy claims rather than reporting unreliable figures, demonstrating the same audit discipline the tool is designed to enforce on the models it evaluates. CeXaR represents a working foundation for trustworthy, context-specific AI assurance in clinical imaging deployment..

Keywords: AI assurance, Audit workbench, Chest X-ray, DICOM, Reproducibility