Radiology AI Testing and Evaluation Resource on AI Medical Imaging Platform (AimSG)

A. Problem Statement (Current State)

Today, more than three-quarters of the 822 FDA approved AI solutions are in Radiology. Local radiology practices have been fully digitised for more than a decade. AimSG was launched in July 2023 to accelerate use of Radiology AI in mainstream practice.

Prior to adoption, there are still questions on the model performance and impact on standard of patient care in actual clinical environments. Today, the approach to pre-deployment testing varies across clinical institutions and may not fully consider all important facets of the model's performance, robustness, governance and generalisability. There is also significant duplication of efforts with comparison of similar models but varying methods and results. There is an unfilled gap in the system where public healthcare institutions (PHIs) can effectively coordinate their efforts in testing AI models before production and deployment. Our aim is to establish Singapore as a leading site for testing and subsequent adoption of Radiology AI models.

B. Challenge Statement

How might we create a standardised and rigorous independent clinical evaluation of AI x Radiology models to drive clinical adoption and confidence in deployment?

C. What Are We Looking For? (To-Be State)

Radiology AI model evaluation requires clinical end-users from PHIs to coordinate the efforts in testing the AI model prior to production and deployment. A standardised evaluation and validation approach leveraging each individual strengths will best serve the public and patients' interests, further grow the burgeoning field of Radiology AI, and sustain user trust in these technologies for medical care provision.

Join us in the search for technology-enabled solutions to achieve the desired state of enhancement for the following:

1. **Performance Metrics:** For example, sensitivity, specificity, accuracy, trade-offs across model threshold intervals, as well as pixel vs image level analysis with suitable uncertainty estimation methods.

2. **Operational metrics** (these will be defined per use case and aligned with clinical experts to determine the intended use of the solution):
    i. Intended clinical deployment setting e.g. primary health care, emergency room.
    ii. Intended use case e.g. diagnosis, triaging.
    iii. Clinical significance of findings.

3. **Dimensions:** Performance metrics should be further broken down by dimensions of analysis such as, but not limited to age, gender, finding type and imaging source.

4. **Robustness Metrics** (these will be defined per use case and aligned with clinical experts):
    i. Common corruptions and perturbations e.g. noise, image compression.
    ii. Evaluating use of external artefacts (to be defined with clinical experts) to supplement data collection and curation.

5. **Technical metrics:** For example, latency, throughput and uptime.

6. **AI Governance** dimensions aligned with Singapore's AI Verify framework and/or other radiology standards.
    i. Integration of vendor-provided information (e.g., model data specifications, security, data governance) to reporting platform.
    ii. Additional examples of evaluations focusing on operational deployment support that can be provided such as definition of ground truth or reference standard, curation of images to create representative population datasets.

7. **Testing process and user interface:**
    i. The above-mentioned technical and governance tests should be run automatically for every new model integrated into AimSG with an automated refresh of the results on a yearly basis. The reference test sets will be developed by PHIs across the healthcare clusters in Singapore.
    ii. Results should be made accessible to Radiologist through the AimSG platform, in the form of a comprehensive report for individual models of interest and a comparison leaderboard to assist selection of the most appropriate model.

8. **Technical IT requirement**:
    i. Integration: The proposed solutions must be able to interface with AimSG's APIs.
    ii. Scalable: The proposed solutions must be easily scalable to allow multiple users from different clusters to use at the same time.
    iii. Well-secured: Any recommended solutions must undergo regular risk assessment and adhere to the cybersecurity standards to secure the data and safeguard the systems.

9. **Scope and Timeline:**

    This collaboration is to establish a standardised and rigorous independent Clinical Evaluation Methodology, Metrices, Guidelines and to Co-develop synthetic artefacts with PHI's users for the following AI x Radiology models:

    i. Chest X-ray AI Model
    ii. Mammography AI Model
    iii. Fracture X-ray AI Model

    The deliverable of the collaboration is to develop a solution with Integration to AimSG pre-deployment platform, which can generate a comprehensive report for individual models of interest and a comparison leaderboard to assist selection of the most appropriate model.

    The total duration of the collaboration should not be more than 1.5 years from the commencement of the collaboration.