Conference On Machine Intelligence In Medical Imaging
SIIM CMIMI24 Oct 21-22 | Boston, MA

Ensuring Real-Time Reliability: An Autonomous Monitoring System for Radiology AI Performance

Suyash Khubchandani, MD, MHA, CARPL.ai, Inc.; Vasantha K. Venugopal, MD; Abhishek Gupta; Rohit Takhar

Introduction/Background

Integrating artificial intelligence (AI) in healthcare, especially radiology, has revolutionized diagnostics. However, maintaining AI model accuracy in real-time clinical settings is challenging, primarily due to the lack of real-time ground truth data. This study introduces an autonomous monitoring system using two novel metrics: predictive divergence and temporal stability, providing real-time insights to ensure AI model reliability.

Methods/Intervention

To overcome real-time monitoring challenges without ground truth data, we developed two key metrics: Predictive Divergence: This metric employs Kullback-Leibler (KL) and Jensen-Shannon (JS) divergences to compare predictions between the primary AI model and two supplementary models. Lower divergence indicates higher accuracy and agreement among models. Temporal Stability: This metric assesses AI model consistency by comparing current predictions with historical moving averages. Variations in temporal stability can indicate model decay or data drift. The system was validated using chest X-ray data from a single-center clinic. Three commercial AI models for chest X-ray classification were analyzed in a longitudinal retrospective study design, using Jensen-Shannon Divergence (JSD) to compute predictive divergence and temporal stability metrics.

Results/Outcome

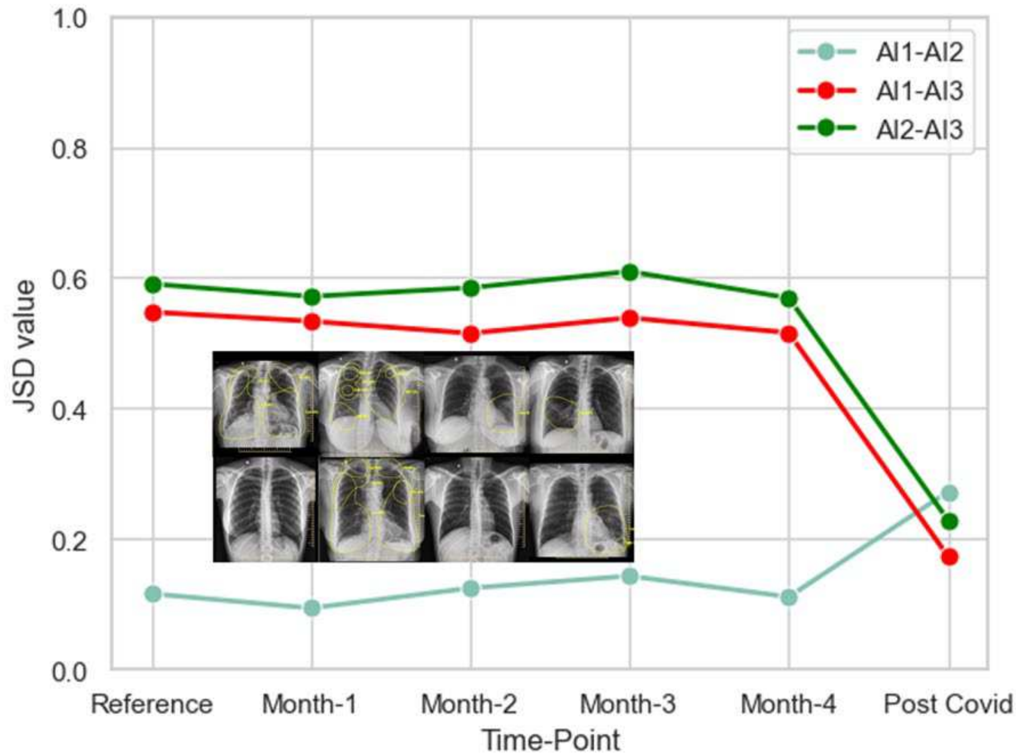
The study analyzed 3,993 chest X-rays over several months, including the onset of the COVID-19 pandemic. Key findings include: Predictive Divergence: JSD values initially showed alignment between the main AI model (AI1) and its support models (AI2, AI3). Post-COVID, a significant increase in divergence between AI1 and AI2 indicated a need for model intervention. Temporal Stability: JSD values for AI1 indicated initial consistency, but increased significantly post-COVID, reflecting a deviation from historical performance. AI2 and AI3 also showed increased divergence post-COVID, highlighting the pandemic's impact on AI model predictions.

Conclusion

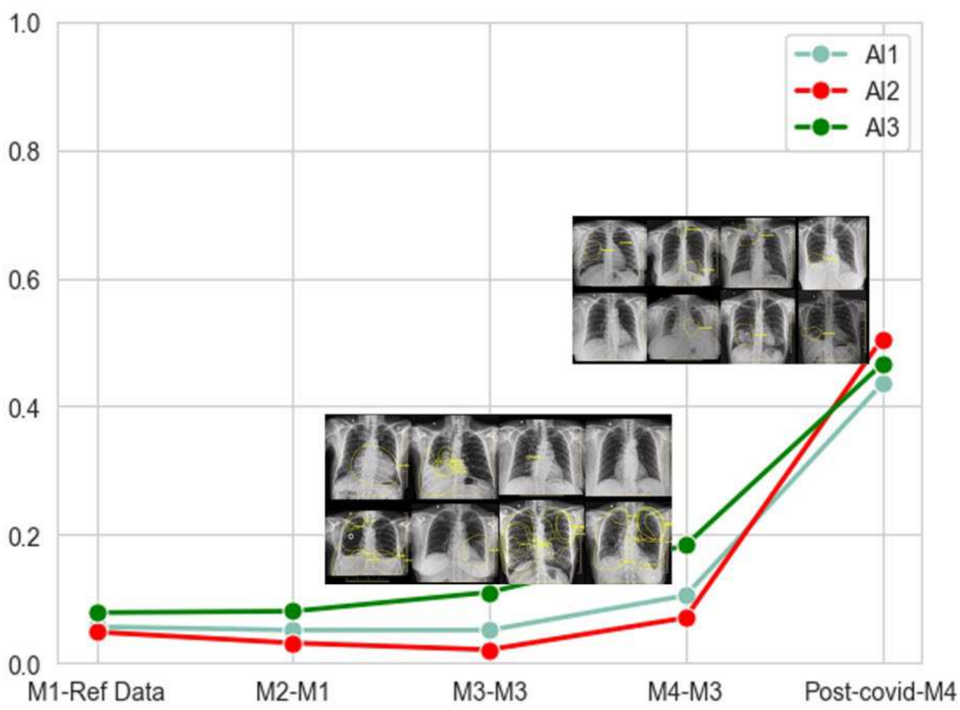
The proposed system, using predictive divergence and temporal stability, offers a robust framework for real-time AI performance evaluation in clinical settings. This ensures the safe integration of AI in healthcare, as demonstrated during the COVID-19 pandemic. The system's continuous insights can enhance AI model reliability, ultimately improving patient care.

Statement of Impact

Continuous AI model monitoring in healthcare is crucial. The proposed metrics enable real-time detection of performance issues without ground truth data, significantly enhancing AI model reliability in clinical practice. Future research will optimize this system for various clinical contexts and AI models.



JSD between main model and support models



Temporal Stability of all three AIs at the monthly check points

Keywords

Deep learning; Post market surveillance; Mlops; AI Monitoring; Drift detection