Reprinted from *Canadian Healthcare Technology,* June/July 2024 issue

CANADIAN
Healthcare Technology
www.canhealth.com

The missing piece in AI adoption: Real-time performance monitoring

BY JEFF VACHON

The radiology software industry estimates that up to 15-25 percent of medical imaging providers have deployed AI image analysis, often as tests and pilot projects. The goal, of course, is to test the AI solution and if successful, graft it onto regular clinical practice.

The cautious uptake of AI is attributed to various factors, most notably the absence of effective methods to monitor AI performance in practical settings. This involves the challenge of accurately measuring algorithm accuracy and performance in real-time clinical practice.

Ensuring the transparency and accuracy of AI algorithms is crucial for fostering trust among clinicians, patients, and regulatory bodies.

It assists the radiologists to better understand and have greater confidence in the AI-driven decision-making process. And as a support tool, AI can lead to more confident and informed decision making.

AI Performance Monitoring should provide clear metrics on the performance impact of all AI algorithms. It should include measures such as accuracy, and any limitations or biases present in the AI's predictions based on the real-world data of the client's own patient demographics.

Tracking factors also include Concordance, Discordance, Sensitivity, and Specificity.

Alongside the immediate benefits of real-time monitoring, longer-term monitoring provides invaluable Drift reporting, the predictive analysis of performance over time. AI metrics should also assess the impact on radiology productivity, determining the effectiveness and efficiencies of AI in a busy radiology practice.

To effectively oversee and enhance AI systems, it is helpful to use a specialized integration platform. A platform enables healthcare organizations to deploy multiple AI algorithms across different clinical subspecialties and operational areas via a single point of contact and contract.

Applications that monitor AI performance will then complement all clinical applications deployed on the platform.


Jeff Vachon

One example of such a platform is the Bialogics system, which resulted from an alliance between the Canadian developer Bialogics Analytics and Scotland-based Blackford Analysis.

"The collaboration between Blackford and Bialogics Analytics creates a real-time feedback system for AI outcome management and performance analytics with the goal to improve clinician confidence and workflows and ultimately patient outcomes," commented Ben Panter, founder and CEO of Blackford.

The platform extracts data from any one of a multitude of AI algorithms and measures output and integrated performance metrics. It does this via a real-time dashboard covering Concordance, Discordance, Sensitivity, Specificity, as well as Drift Reporting, enabling accurate performance analysis and productivity predictions.

Is your radiology organization AI data-ready? Preparing a radiology organization for AI data readiness, involves the establishment of an integrated system that is vendor-agnostic and adaptable to AI technologies. This system must be capable of aggregating data from multiple Radiology Information Systems (RIS) and Picture Archiving and Communication Systems (PACS), as well as outputs from various AI algorithms and the clinical findings by radiologists.

The output of the AI algorithms, essentially the AI algorithm findings, is often collected using secondary capture and DICOM overlay techniques and typically stored within the DICOM header.

This output when extracted is then compared to the previously collected operational and clinical data and the result of this comparison comprises AI Performance Monitoring. The integration and analysis of data from many diverse sources is critical to AI Performance Monitoring.

The operational data from the radiology workflow is integrated with the clinical data which is then extracted and analyzed by machine learning (ML) engines. These analyze the text of the radiology report using Natural Language Processing (NLP).

AI performance monitoring metrics: There are five primary operational analytics that need to be measured to provide real-time monitoring of AI in the radiology environment.

1. Concordance/Discordance Analysis: Concordance and Discordance reporting within radiology AI involves contrasting the diagnoses or assessments generated by the AI system against those determined by radiologists, often regarded as the 'ground truth'. It's a straightforward assessment that entails comparing the AI's binary output with the findings of the radiologist. Concordance being the number of times both AI and radiology findings are concurrent, and Discordance being the number of times findings are not concurrent.

2. Sensitivity and Specificity: These measurements play a vital role in evaluating the diagnostic accuracy of AI algorithms in radiology. A balance between sensitivity and specificity is essential.

• Sensitivity: Indicates the proportion of actual positive cases correctly identified by the AI system. It measures the ability of algorithms to correctly identify individuals with a certain condition.

• Specificity: Measures the proportion of actual negative cases correctly identified by the AI system. It measures the ability of the algorithm to correctly identify individuals who do not have a certain condition.

3. AI Confidence Scoring: In radiology, this refers to the algorithm's assessment or estimation of its own reliability or certainty regarding a particular diagnosis or decision made based on the imaging data.

When an AI system analyzes medical images, it doesn't just provide a diagnosis but often assigns a level of confidence or certainty to that diagnosis.

However, not all AI algorithms calculate this value. Confidence scoring involves the AI algorithm assigning a probability or confidence level to its output, indicating how confident it is about the accuracy of its assessment.

The cautious uptake of AI is attributed to various factors, most notably the absence of effective methods to monitor AI performance in practical settings.

4. AI Impact on Radiology Efficiency: Using pre- and post-AI evaluation times. For example, Radiologist Turn Around Times (TAT), Read Times and Productivity measurements like Relative Value Units (RVU), all of which combine the messages from HL7, DICOM, look up tables and the conversion of unstructured textual data to structured data from the clinical report using Natural Language Processing can create;.

• Efficiency measurements assist in allocating resources effectively. Identifying which cases or modalities benefit the most from AI assistance helps in directing resources and AI tools to areas where they can make the most significant impact.

• Effectiveness monitoring enables the comparison of radiologist read times before and after AI implementation for the assessment of AI's effectiveness. If there's a notable reduction in read times without compromising diagnostic accuracy, it showcases the productivity gains achieved through AI integration.

5. AI Drift Reporting and Predictive Trending: AI Drift refers to the gradual change or degradation in the performance of an AI algorithm over time due to various factors such as changes in data patterns, evolving patient demographics, or changes in modality configurations.

Drift reporting involves the continuous assessment and analysis of AI algorithm performance against a set of predefined benchmarks or standards.

Any deviations or discrepancies in the algorithm's performance from the expected standards are either identified in real-time or more likely over a period of time, to see progressive changes in the number of discrepancies, or changes to key metrics being used to measure the effectiveness of the algorithm.

Predictive trending involves forecasting future patterns or trends based on historical data and ongoing observations from AI algorithms.

Jeff Vachon is the Founder and President of Bialogics Analytics.