Artificial Intelligence Solutions in Radiology Clinical Practice

Authors: Darcy Richards, Editorial Assistant

ARTIFICIAL intelligence (AI) was a hot topic at the 33rd annual European Congress of Radiology (ECR) 2023, held in Vienna, Austria, between the 1st–5th March. A fascinating open forum session, chaired by Martin Reim, Tartu University Hospital, Estonia, and Chairperson of the European Society for Radiology (ESR) Education Committee Radiology Trainees Forum, was held on the first day. The session, entitled 'Artificial intelligence: questions you wanted to ask us but didn't', included presentations from three experts, and focused on key questions trainees might have about AI in the field of radiology. The session covered AI performance metrics, how AI can be implemented into clinical practice, and how to start-up an AI solution business.

DEFINING ARTIFICIAL INTELLIGENCE AND ITS USE IN RADIOLOGY

Vicky Goh, Department of Radiology, Guy's and St Thomas' NHS Foundation Trust, London, UK, explained that AI is the ability of technology to perform tasks commonly associated with intelligent beings. Goh briefly commented on the evolution of AI and the increasing interest and implementation of AI in healthcare, highlighting that automated acquisition to improve standardisation; post-process deep learning reconstruction to improve image quality; and interpretation with deep learning detection on scans as the main areas that AI currently impacts radiology practice and workflows.

Goh discussed the importance of understanding algorithm performance, and explored the different performance metrics radiologists should be aware of when assessing AI algorithms, including the dice similarity coefficient as a measure of segmentation performance; plus binary classification performance metrics, including binary classification thresholds to calculate true and false positive and negatives, sensitivity and specificity, positive and negative predictive values, accuracy, and F1 scores as an average of precision and recall. Performance metrics that are independent of threshold values, including receiver operating characteristic curve and area under the receiver operating characteristic curve as a measure of how well the algorithm performs classification; precision recall curve and area under the precision recall curve as a measure of how well the algorithm performs recall and precision; and the impact prevalence has on these measures, were also covered.

The full potential of AI in radiology has yet to be achieved, and the market for AI solutions in healthcare is expanding. As such, Goh stated that having an understanding of AI and its implications is important for practising radiologists. In an emerging market, guidance on assessing AI software will be necessary to aid clinicians in decision-making when procuring these for their departments.

ASSESSMENT FRAMEWORK FOR CONSIDERING PROCUREMENT OF ARTIFICIAL INTELLIGENCE SOFTWARE

Patrick Omoumi, Department of Radiology, Lausanne University Hospital (CHUV), Switzerland, delivered an insightful presentation on a framework for how to implement commercial AI solutions into clinical practice using the

evaluating commercial AI solutions in radiology (ECLAIR) guidelines. This framework considers five key sectors for evaluating and implementing commercial AI solutions: relevance; performance and validation; usability and integration; regulatory and legal aspects; and financial and support services.

> **"Having an understanding of AI and its implications is important for practising radiologists."**

In terms of relevance, Omoumi explained that AI solutions must solve a problem, and clinicians should consider how the software would be used, the desired output, and impact on all stakeholders, keeping the anticipated end-user in mind.

With respect to performance and validation, Omoumi highlighted the importance of diagnostic accuracy guidelines for algorithm validation. Examples of these guidelines include the Standards for Reporting of Diagnostic Accuracy Study-AI (STARD-AI) and Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis-AI (TRIPOD-AI), which look at how algorithms are trained; the dataset used to train them and whether

this is representative of the population it will be used on; how performance was evaluated; and whether potential sources of bias have been accounted for.

In addition to relevance and performance, clinicians should also consider usability and integration when evaluating AI software. Omoumi commented an even if an algorithm performs well, it needs to integrate with the institution's workflow to be usable. Utilising a platform that orchestrates different types of algorithms, such as an app store, is a way to overcome institutional workload associated with integrating algorithms into workflows. Omoumi also discussed privacy concerns associated with cloud-based algorithm integration strategies, spotlighting that some institutions do not allow cloud-based algorithms despite the fact that these are easier to integrate and have better access to platform support.

Other crucial elements in algorithm assessment are regulatory, legal, and financial aspects, with regulatory bodies such as the U.S. Food and Drug Administration (FDA) and Conformité Européenne (CE) providing clearance for AI algorithms. Omoumi advised that this should be checked prior to purchasing AI software.

Pricing will be a critical point for many healthcare

services, in which budgets and financial resources are stretched or constrained. Omoumi warned that given that AI solutions in healthcare is an emerging market, valuations are unstable, and vendors may increase license fees year on year, or remove their products from the market unexpectedly. As a result of this, clinicians will have to determine how much to pay for these algorithms. Omoumi discussed that performing a return-on-investment assessment is an effective method to assist in this decision-making. However, this will pose its own challenges, because the benefits of an algorithm may not be financial, but rather improved efficiency or work quality. Ongoing work to aid clinicians and institutions in determining how to pay for algorithm software will be needed moving forwards.

BENEFITS AND CHALLENGES IN IMPLEMENTING INTO CLINICAL PRACTICE

Whilst AI technology has several potential benefits for patients, healthcare professionals, and medical institutions, inevitably there are also some associated risks, both of which were explored throughout the sessions.

Goh touched on the implications of AI for practicing radiologists, spotlighting that AI may be beneficial for overcoming human error, for ensuring quality control that can be difficult in busy day-to-day practice and with a reduced

work force, as well as for helping with capacity issues and workload burden. Moreover, Goh stated that AI-automated image diagnosis is projected to potentially yield annual savings of 3 billion USD by 2026.

Whilst detailing the ECLAIR framework, Omoumi discussed that the risks and benefits to all stakeholders implicated in workflows, from patients to the institution, should be considered. Benefits including increased efficiency or quality were covered, alongside risks such as missed diagnoses or misdiagnosis, given that AI solutions are currently narrow and task-specific.

Goh highlighted that one of the major benefits of AI in radiology is its ability for complex pattern recognition in large sets of data that human operators may miss, with the ability to subsequently convert these patterns into a quantitative format. Furthermore, all three experts considered how AI algorithms, such as machine learning, have the ability to improve their performance over time as they become exposed to increasing volumes of data, with Goh commenting that for specific tasks, machines are "almost achieving human-like performance" and may exceed human performance in the future.

However, there are challenges in assessing an algorithm's performance within an institution's dataset. Omoumi and Goh both discussed how AI software performance is often assessed in enriched populations, and that testing within
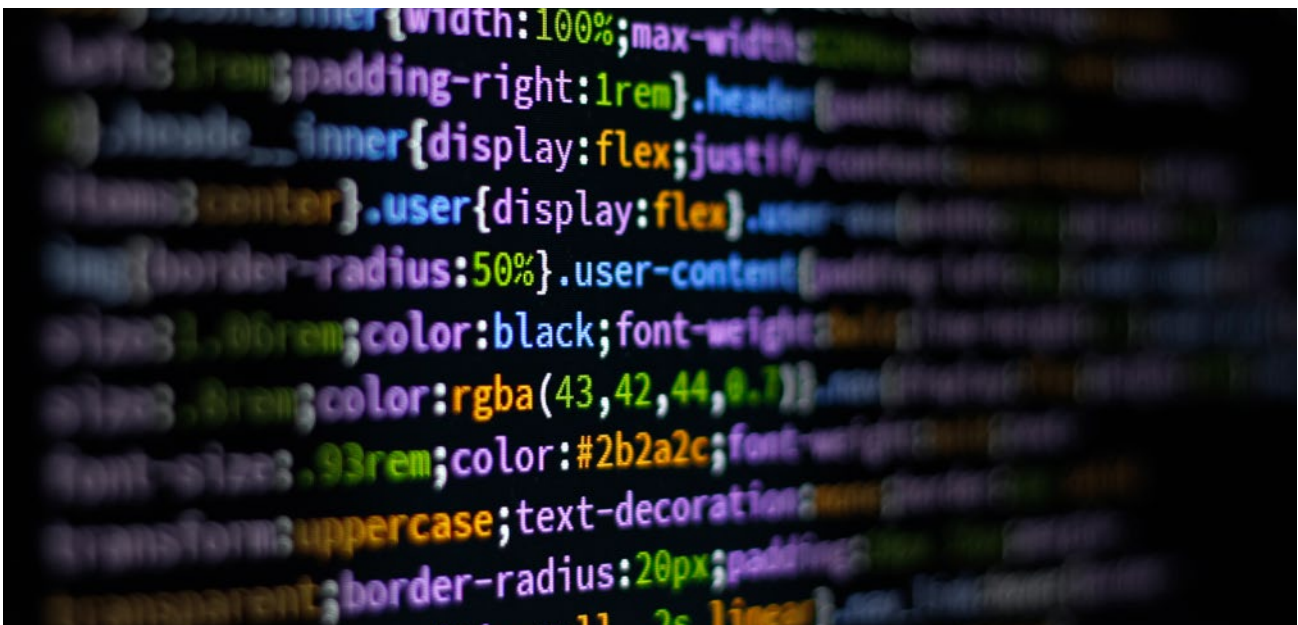
an individual institution's dataset often shows significantly reduced performance compared to literature-reported values. Omoumi spotlighted that performance could be markedly reduced in 25% of cases.

In addition to this, there are concerns regarding the generalisability of AI solutions across different datasets. Omoumi discussed that theoretically, an appropriately developed AI solution should perform reasonably well across different datasets, and further added that some algorithms are capable of continuous learning, and therefore able to improve their performance over time. They highlighted the importance of understanding the software limitations within an institution's dataset, and that currently, there is no way to know this information unless you perform the study yourself.

Another risk in using AI solutions is the potential negative impact on trainee education, and a reduction in basic imaging science knowledge through software over-reliance. Alongside this, there is the potential for negative impact on workflows; for example, a tool with high sensitivity and low specificity would yield a high false positive rate, increasing workload.

Validation in image segmentation was also discussed as a challenge. Goh said that noise in standalone reference data used to train algorithms due to inter-reader variability is a key concern with segmentation training. The use of algorithms capable of continual background learning alongside experts changing their image annotations in real-time as a method to perfect the segmentation algorithm could be a potential solution. However, in practice, this may be unrealistic and impractical.

CONCLUSION

AI in radiology is an exciting and emerging field, with significant potential to improve practitioner workflows and patient experiences; however, there are several concerns and challenges that will need to be addressed over the coming years. Reim closed the open forum by inviting a continued offline conversation about the topics raised throughout the session. ●

> "In addition to relevance and performance, clinicians should also consider usability and integration when evaluating AI software."

 EMJ