

AI: Opportunities, Capabilities and Limits

THE JOURNAL 2022

Henrique Martins et al.
Hospitals-on-FHIR: Preparing Hospitals for European Health Data Space

Rafael Vidal-Perez
Artificial Intelligence and Echocardiography: Are We Ready for Automation?

Konstantinos Petsios et al.
Artificial Intelligence in Radiology: Realities, Challenges and Perspectives from a Tertiary Cardiac Centre in Greece

Sai Pavan Kumar Veeranki et al.
Learning From Each Other: An Artificial Intelligence Perspective in Healthcare

Elmar Kotter
Integrating Decision Support and AI in Radiology

†Werner Leodolter
Clinical Decision Support – Benefits and Application in Healthcare





The Current State of AI in Diagnostic Imaging and How to Improve its Clinical Value

Thomas Kau | Head | Department of Radiology | LKH Villach | Villach, Austria

More well-curated data, external validation of algorithms, and user-friendly workflow integration will help translate the potential of AI into clinical routine.



Key Points

- Representative datasets and external validation are key issues for AI algorithms to become more accurate and generalisable.
- Machine learning is not only used for automation; it will also help expand the vocabulary of patterns that lead to treatment decisions or find new relationships.
- Although individual tasks are taken over by AI, computers will augment human decision-making rather than replace imaging professionals.

Although an ever-increasing number of publications suggests that artificial intelligence (AI) could provide value in numerous medical imaging applications, there seems to be a considerable gap in moving from proof-of-concept to production. Once a software solution has made it to commercialisation, it still needs to prove its robustness and deliver its benefits within a well-established clinical workflow. To date, only a limited number of AI solutions provide evidence of meeting the ultimate goal of any healthcare technology, which is to improve patient outcomes. By posing four provocative questions, this article suggests how to enhance the value of AI for radiology and other diagnostic specialities.

Are Commercial AI Solutions Accurate and Generalisable?

Performance consistency is increasingly becoming the focus of scientific research. In their recent exhaustive review, Brendan Kelly and co-workers found a propensity for bias and a lack of generalisability for many published algorithms (Kelly et al. 2022). Across all studies included, the median performance for the most utilised metrics was a Dice score of 0.89, an area under the curve (AUC) of 0.90, and an accuracy of 89.4. Average performance in those studies allowing for direct comparison decreased by 6% at external validation, ranging from an increase of 4% to a decrease of 44%. The requirement for external validation of AI solutions even after approval

by regulatory authorities corresponds to published evidence including our own experience derived from a neuro-computed tomography (CT) study (Kitamura et al. 2021; Kau et al. 2022).

If a decision support software produces a limited number of erroneous alerts which experienced physicians can easily resolve, it may still hold promise for triage and notification (Figure 1). Statistics are critical in assessing AI performance. Therefore, it should be no surprise that metrics are used to the advantage of the respective product, at least in marketing brochures. The AUC provides a single aggregated measure. However, Elad Walach, CEO of Aidoc, once explained that it is confusing to physicians and may be overemphasised (Walach 2019). The negative predictive value of a decision support software should be really high, while a relatively low positive predictive value could still be satisfying for detecting rarer diseases. Ideally, disease prevalence in a training data set should correspond to the real-world clinical scenario. Apart from the size of a convolutional neural network, diverse and meticulously labelled data sets and better disease models are key issues for AI algorithms to become more accurate.

Do Current AI Developments Reflect Clinical Needs?

Much of the recent excitement surrounding AI in diagnostic imaging gravitates toward algorithms developed for interpretative tasks. In fact, AI promises to have an impact on

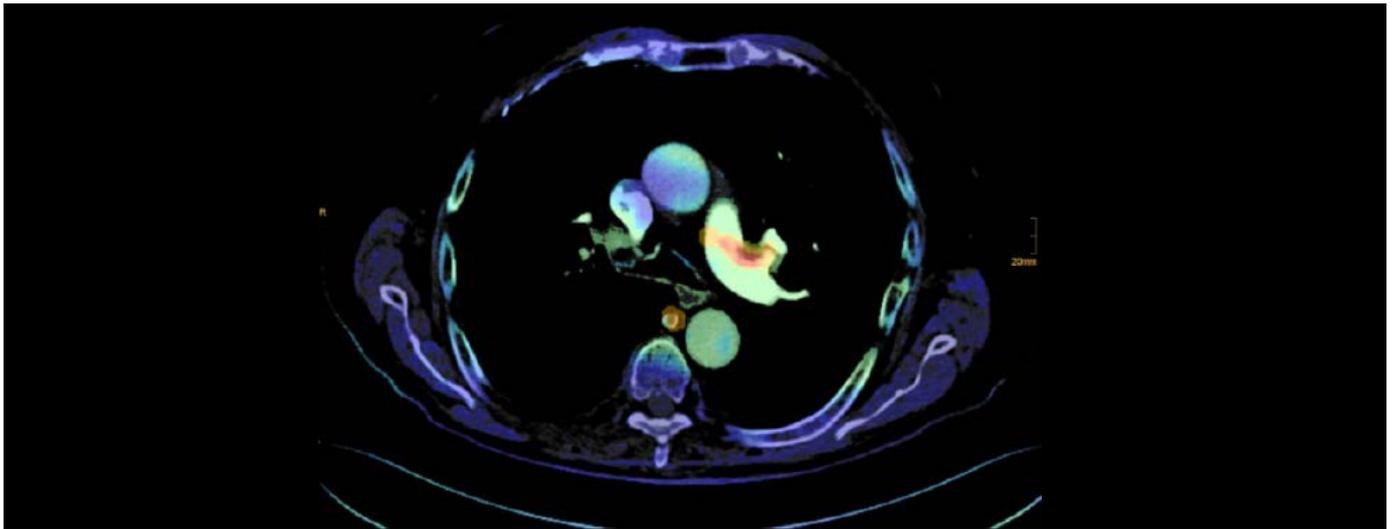


Figure 1: Contrast-enhanced thoracic CT analysed by a commercial triage and notification software that flags and communicates pulmonary embolism. While correctly marking an embolus in the left pulmonary artery (*red, anteriorly*), it erroneously also highlights the azygos vein (*red, posteriorly*).

medical diagnostics along its entire value chain (Richardson et al. 2021). That ranges from using deep learning towards dose optimisation to case prioritisation and from automated segmentation tasks to pattern recognition. To develop an additive value in clinical routine, isolated solutions must be integrated into an efficient workflow. With different degrees of maturity, marketing framework, and media attention, we see AI developments focussing on patient safety, image quality, reduction of reading time, or decision support. A promising research area, yet slowly translating into practice, is the characterisation of “radiomic” features, which are not discernible by visual inspection (Bera et al. 2022). Such MRI fingerprints have been shown to be predictive of treatment response in specific diseases, which may even lead to the re-classification of certain pathologies. Not without reason, AI is considered a key tool on the road to personalised precision medicine.

It is doubtful for various reasons, not least regarding data-sharing and economy, whether an undisputed wish of healthcare professionals will soon be taken into focus, namely the development of decision support algorithms for rare diseases. Think, for example, of pattern recognition in acquired and, moreover, congenital metabolic brain disease. Its aetiopathogenetic diversity with sometimes similar imaging patterns, gradual as well as inhomogeneous expression, and infrequent appearance of single types make this neuroradiologic topic a traditional challenge for differential diagnostics. In current commercial software, there is a clear tendency towards solutions for frequent diseases or imaging requests, respectively.

Is There a Risk That AI Will Deskill or Even Replace Radiologists Instead of Complementing Them?

In many countries around the globe, as the demand continues to increase, the emerging support offered by disruptive

technologies meets a current supply crisis in diagnostic imaging services. It is expected that algorithms will overtake some of the core tasks of radiologists. Whenever the issue is narrowed down to the question of whether humans or machines produce better results, we should not forget that AI systems work with high levels of standardisation and without being vulnerable to fatigue or cognitive biases – while being prone to other sources of bias. At the moment, however, we are dealing almost exclusively with single task solutions which, if sufficiently accurate, take on individual tasks such as the detection of pulmonary embolism. By producing timely alerts, they steer the radiologist’s eye in the right direction. On the other hand, we are currently far from AI software covering the entire differential diagnostic spectrum, especially in complex clinical scenarios. Needless to say, several expectations are linked to the patient/doctor relationship of trust, ranging from contextual experience via personal responsibility to medical advice. The level of additional diagnostic accuracy, consistency, and, ultimately, patient outcomes will likely depend on the careful coordination of machine and human capabilities.

Breast imaging is a very promising use case for AI as it is expected to impact several of the aforementioned aspects. As for screening mammography, recent evidence (McKinney et al. 2020) suggests that AI algorithms may soon be able to replace the second reader for unequivocal cases, saving resources to a limited extent. However, there is a lack of compelling evidence from large prospective studies based on models fusing AI and human expert vision (Lehman et al. 2021).

Generally, in the foreseeable future, computers will augment human decision-making instead of replacing imaging professionals. While triggering a need for education in basic principles of machine learning methods, the introduction of AI software also includes the risk of deskilling among diagnostic imaging



specialists, as put up for discussion by Michael Fuchsjäger in the course of last year's AICI Forum (Leodolter 2021). In order to be able to check the plausibility of software outputs, future radiologists, pathologists, and other diagnostic specialists will still need classic knowledge and skills. Furthermore, they are

training and testing of algorithms. Such requirements challenge the legal and ethical framework towards enabling secure data-sharing for the public good and, in patient care, cloud-computing versus on-premise solutions. Once deployed in a radiological workflow, AI software is expected to perform

AI systems work with high levels of standardisation and without being vulnerable to fatigue or cognitive biases – while being prone to other sources of bias

expected to keep a holistic approach in the increasingly overlapping field of medical diagnostics. Finally, augmented teaching tools will expand medical training in a meaningful way.

What are the Challenges for AI to Provide Value in Clinical Routine?

Translating the potential demonstrated in AI research into a measurable clinical benefit requires healthcare stakeholders to be aware of all the expectations, challenges and pitfalls that can accompany the development and deployment of AI systems in a diagnostic imaging environment. In his recent article on “Separating hope from hype” in AI for radiology, Jared Dunnmon stressed the importance of meaningful performance measurement for a consistently defined and clinically relevant task (Dunnmon 2021). To deliver repeatable and reproducible results, datasets used for evaluation must represent the characteristics of a population on which an algorithm is intended to be applied, including sufficient granularity to capture relevant variations. This leads to the perception of AI algorithms as a “black box”. The higher the rate of false-positive and even more so false-negative results, the greater is the demand for interpretability. Further research is warranted on explainable AI, although explainability is not an exclusive parameter of trustworthiness and certainly not a good indicator of accuracy.

Another hurdle for translation into clinical care is the lack of large, diverse, and well-curated data sets for independent

consistently over the full range of possible imaging patterns and present results timely on user-friendly interfaces. For the time being, healthcare professionals are still “sipping from the glass”, just having a try with a few algorithms here and there. In the near future, the industry will likely provide something like a marketplace or app store for imaging specialists to choose from a preselected number of AI tools which may be always on or work on click (Saboury et al. 2021).

It is beyond the scope of this article to elaborate on economic concerns for healthcare providers such as budgeting and reimbursement of AI supply. What is becoming more and more apparent, though, are heavily hardware-centric budget plans confronted with a dynamically growing software market.

In conclusion, AI holds the potential to provide value by improving patient access to expert service, extracting information and new knowledge from imaging data, increasing diagnostic certainty, saving time, and ideally improving clinical outcomes. In order to get disruptive changes realised beyond the hype, several stakeholders should keep fusing their ideas and viewpoints on all available levels. One of those opportunities will be the 4th AICI Forum taking place in Graz, Austria, this autumn.

Conflict of Interest

None. ■

REFERENCES

Bera K et al. [2022] Predicting Cancer Outcomes with Radiomics and Artificial Intelligence in Radiology. *Nat Rev Clin Oncol*. 19:132-146.

Diagnostic Image Analysis Group [2022] AI for Radiology. Available from grand-challenge.org/aiforradiology.

Dunnmon J [2021] Separating Hope from Hype: Artificial Intelligence Pitfalls and Challenges in Radiology. *Radiol Clin North Am*. 59:1063-1074.

Kau T et al. [2022] FDA-Approved Deep Learning Software Application Versus Radiologists with Different levels of Expertise: Detection of Intracranial Hemorrhage in a Retrospective Single-Center Study. *Neuroradiology*. 64:981-990.

Kelly BS et al. [2022] Radiology Artificial Intelligence: A Systematic Review and Evaluation of Methods (RAISE). *Eur Radiol*. Epub ahead of print.

Kitamura FC et al. [2021] Clinical Artificial Intelligence Applications in Radiology: Neuro. *Radiol*

Clin North Am, 59:1003-1012.

Lehman CD, Topol EJ [2021] Readiness for Mammography and Artificial Intelligence. *Lancet*. 398:1867.

Leodolter W [Moderator] [2021] Commercial AI solutions – To Buy or Not to Buy? Panel discussion at 3rd AICI Forum, Klagenfurt, 5-6 Nov. Unpublished

McKinney SM et al. [2020] International Evaluation of an AI system for Breast Cancer Screening. *Nature*. 577:89-94.

Richardson ML et al. [2021] Noninterpretive Uses of Artificial Intelligence in Radiology. *Acad Radiol*. 28:1225-1235.

Saboury B, Morris M, Siegel E [2021] Future Directions in Artificial Intelligence. *Radiol Clin North Am*. 59:1085-1095.

Walach E [2019] Lies, Damned Lies and AI Statistics. Available from medcitynews.com/2019/08/lies-damned-lies-and-ai-statistics/