EDITORIAL COMMENT

Machine learning solutions in radiology: does the emperor have no clothes?

Renato Cuocolo[1,2] · Massimo Imbriaco[3]

Key Points

- *Interest in radiomics and machine learning is steadily increasing and is reflected both in research output and number of commercially available solutions.*
- *Currently available commercial products using machine learning are often supported by limited evidence of clinical usefulness and studies are often of low methodological quality.*
- *Ethical and regulatory issues remain open and hinder implementation of machine learning software packages in daily clinical practice.*

Interest in radiomics and machine learning (ML) has grown exponentially over the years within the field of medical imaging, as demonstrated by the increasing number of recently published papers [1]. As reported in the current issue of *European Radiology*, this trend is also reflected in the number of commercial products using ML that are being offered to radiology units. The authors have presented several issues which may negatively impact the real-world applicability of such software solutions, as well as provide a useful database to guide radiologists in their assessment (www.aiforradiology.com) [2]. Recently, guidelines to assess commercial ML solutions (the ECLAIR guidelines) have also been proposed to help physicians in this task [3]. Among the reported findings, some are particularly worrisome and

may be the result of deeper issues within ML and radiomics research in healthcare.

First, there is the question of scientific validation of CE-marked ML packages which is lacking at best. As clearly shown by the authors in the results section, even though already 100 CE-marked AI products are commercially available, only 36 of these products have peer-reviewed evidence and most demonstrate only lower levels of efficacy [2]. Studies were almost exclusively retrospective and did not sufficiently address clinical impact of the ML software, limiting themselves in most cases to feasibility investigations and isolated performance assessment. This finding is in line with the characteristics of radiomics and ML research in general, as clearly shown by recent systematic reviews using the Radiomics Quality Score (RQS) to evaluate the methodological quality of such investigations. Independently of the application area (e.g., lung cancer, renal cell carcinoma), the average RQS is usually < 25% and often < 10% [4, 5]. The reasons for this underwhelming situation will sound familiar, i.e., lack of prospective design, and lack of robust validation of features, lack of strong model validation, especially in a multi-institutional setting. While these limitations can be partly justified by the need to build up a scientific background to justify future prospective trials, this does not apply to commercially available products which have also received approval for clinical use.

The question of best implementation practices for these software packages within the information technology

✉ Renato Cuocolo
   renato.cuocolo@unina.it

1   Department of Clinical Medicine and Surgery, University of Naples "Federico II", via Pansini 5, 80131 Naples, Italy

2   Augmented Reality for Health Monitoring Laboratory (ARHeMLab), Department of Electrical Engineering and Information Technology, University of Naples "Federico II", Naples, Italy

3   Department of Advanced Biomedical Sciences, University of Naples "Federico II", Naples, Italy

framework currently available in hospitals is also open. As reported by the Authors, there is still some uncertainty regarding the best deployment strategies, and vendors often offer multiple solutions that may be challenging to navigate for the physicians. While the current situation represents an opportunity for companies to act as a middleman and attenuate the issue, this may also introduce biases in the availability of ML solutions for radiologists. A regulated marketplace may ease accessibility for the user but raise barriers for new players with innovative solutions, especially if represented by smaller companies. This trend has been seen elsewhere in software and the trade-off must be known and accepted by physicians [6]. An interesting, alternative solution to this issue could be represented by the implementation of a vendor-neutral infrastructure, as recently proposed by Leiner et al [7].

There are unresolved ethical and regulatory issues tied to the clinical use of any ML package. For example, ethical questions that are now emerging in other image-based ML fields will have to be addressed in healthcare as well. In the domain of facial-recognition software, consent of data use represents a significant problem and is often glossed over by companies offering commercial solutions. This has led recently to heated discussions within the research community and multiple paper retractions [8]. As physicians, we are well-aware of the need for patient and ethics board approval for any study, retrospective or prospective. However, consent must be specific for the research proposal, and using data from a clinical trial for the development of commercial solutions may be a challenging area to navigate. Especially in light of current European Union regulations and attention to privacy, there can be no doubts on the consent of subjects whose exams are used to train ML models which are then included in for-profit commercial products. Companies must be transparent on this point, and no less than full disclosure should be expected from end-users. Regarding European Union citizens, it must also be noted that the General Data Protection Regulation allows the right to erasure based on simple withdrawal of consent from the subject [9]. The implications of this possibility should be considered when purchasing ML software, as the training data could be modified ex post and model performance could change consequently. Also, contrary to the USA, in Europe there, there is no public database available to verify ML software certification or the clinical validation on which it was based. This is compounded by the problem of product certifications in different classes according to its risk. This discrepancy among classes will be overcome and hopefully simplified when the Medical Device Regulation will definitively replace the current Medical Device Directory. Finally, both in the USA and Europe, liability when using ML in healthcare remains an open question from a legislative point of view. This also opens the door to a lack of clear insurance coverage in cases of damage caused by erroneous predictions by ML, limiting the practical usefulness of such software overall [10].

In conclusion, the reviews recently conducted on both scientific studies as well as commercial products prove that radiomics and ML are still in their infancy for healthcare and often not yet ready for use in daily clinical practice. Current buyers must be aware of the current state-of-the-art and especially of the numerous, still unresolved ethical and validation concerns. They should understand that they are early adopters, potentially on the cutting edge of technology but also in part beta testers for these new solutions.

Declarations

References

1. Pinto dos Santos D, Dietzel M, Baessler B (2021) A decade of radiomics research: are images really data or just patterns in the noise? Eur Radiol 31:1–4. https://doi.org/10.1007/s00330-020-07108-w
2. van Leeuwen KG, Schalekamp S, Rutten MJ et al (2021) Artificial intelligence in Radiology; 100 commercially available products and their scientific evidence. Eur Radiol. https://doi.org/10.1007/s00330-021-07892-z
3. Omoumi P, Ducarouge A, Tournier A et al (2021) To buy or not to buy—evaluating commercial AI solutions in radiology (the ECLAIR guidelines). Eur Radiol. https://doi.org/10.1007/s00330-020-07684-x
4. Chetan MR, Gleeson FV (2021) Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. Eur Radiol 31:1049–1058. https://doi.org/10.1007/s00330-020-07141-9
5. Ursprung S, Beer L, Bruining A et al (2020) Radiomics of computed tomography and magnetic resonance imaging in renal cell

carcinoma—a systematic review and meta-analysis. Eur Radiol. https://doi.org/10.1007/s00330-020-06666-3

6. Mac Sithigh D (2013) App law within: rights and regulation in the smartphone age. Int J Law Inf Technol 21:154–186. https://doi.org/10.1093/ijlit/eat002

7. Leiner T, Bennink E, Mol CP et al (2021) Bringing AI to the clinic: blueprint for a vendor-neutral AI deployment infrastructure. Insights Imaging 12:11. https://doi.org/10.1186/s13244-020-00931-1

8. Van Noorden R (2020) The ethical questions that haunt facial-recognition research. Nature 587:354–358. https://doi.org/10.1038/d41586-020-03187-3

9. (2019) EU General Data Protection Regulation (GDPR), third edition. IT Governance Publishing

10. Gerke S, Minssen T, Cohen G (2020) Ethical and legal challenges of artificial intelligence-driven healthcare. In: Artificial Intelligence in Healthcare. Elsevier, pp 295–336