Physica Medica 98 (2022) 8–10

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Editorial

Artificial intelligence applied to medicine: There is an "elephant in the room"

"…..AI algorithms used for diagnosis and prognosis must be explainable and must not rely on a black box…..".

This strong and courageous sentence recently captured our attention: it was stated by S Kundu, a highly reputed young scientist, in her short editorial/letter recently published on Nature Medicine [1]. Dr Kundu has a very intriguing profile being both a physician and a computer/engineering scientist: despite her young age, she received several recognitions being named as "one of Forbes 30 under 30, MIT Technology Review's 35 innovators under 35, a World Economic Forum Global Shaper and a winner of the Carnegie Science Award" as reported in her Wikipedia page [2]. Thanks to her rapid career, a broader public increasingly knows her, becoming an ambassador of the need for transparent AI in medicine. She spoke at the recent United Nations AI for Good global summit and inspired one of its priority areas of sustainable development Goals. She is a physician and a researcher at the Department of Radiology of Johns Hopkins University (JHU) in Baltimore, not by chance, mostly working on AI applications in imaging. AI applications to medical images currently represent a paradigmatic example of the "lights" and "shadows" of the introduction of AI in clinical environments. The Dr Kundu's editorial gave us the opportunity to discuss several aspects that stay at the core of the actual process of AI introduction in medicine, including visions concerning how medical physicists may actively contribute within this complex scenario.

The lack of explainability/interpretability: An "elephant in the room"

Consider a "well trained" AI-based model (model 1) able to estimate the probability that a certain patient affected by a specific type of tumor will not respond to a particular therapy. Suppose that the discriminative power of such a model (incorporating the "hidden" contribution of tens or hundreds of variables) has a declared (after some serious independent validation) area under the curve (AUC) of 0.95. On the other side, for the same clinical situation, an alternative predictive model (model 2), including only a few features/parameters whose meaning is well interpretable or at least understandable, may offer the same prediction with an AUC of around 0.85. Is there someone who doubts what model will be preferentially chosen if the dilemma is proposed to a large group of expert doctors? We may guess that the majority would choose model 2. This simple and hypothetical example well describes clinicians' absolute preference to the possibility of "understanding" why the model they are using works. Despite this evidence, the image of a medicine more and more driven by AI tools that support (and maybe in the future partly replace) the doctor in making decisions concerning the care of patients is largely promoted in the scientific world and still more in the media and the public. We may argue how many interests and pressures from industry and technological players may act at various levels to push ahead the penetration of AI in the medical field.

However, the lack of explainability of black-box AI tools emerged in the recent literature as a relevant issue [3–7] and is increasingly debated at different levels, including policymakers, ethics experts and philosophers.

Notably, the European General Data Protection Regulation (GDPR) [8] includes a first fundamental step, pointing attention to this issue and including the right for an individual to obtain a meaningful explanation when automated (algorithmic) decision-making is involved. This includes the information on the existence of automated decision-making, meaningful explanation of the logic involved, the significance and envisaged consequences of such processes.

In our opinion, the recent emphasis on the road toward "explainable AI" [9–11] should be considered as a first attempt to recognize the presence of the "elephant in the room". On the other hand, we are far from having the solution: the elephant is there, and it is hard to push it out of the room!

We propose some considerations, focusing on the position of Medical Physics within Medicine and, likely more, in the traditional domains where medical physicists are mainly asked to deal with AI. Hopefully, this will contribute to better orienteering us in the current initial process of AI implementation.

AI applications with "intrinsic usability": Moving the elephant out

It is out of doubt that several applications of AI may be considered as intrinsically explainable or at least interpretable or usable. Not by chance, they represent areas of likely more rapid implementation of AI-assisted tools that promise to improve quality and efficiency in different domains of medicine. A paradigmatic example is the AI-based automatic segmentation for radiotherapy applications [12] and, more and more, in imaging applications due to the growing use of quantitative information during the diagnostic process. We can also mention the ever-increasing applications of AI in image reconstruction, aimed to obtain images with better contrast and resolution [13] or the spread of AI-based approaches for radiotherapy plan optimization, shifting an increasing fraction of manual procedures toward automation [14,15]. These are only a few and not exhaustive examples of AI implementation, often largely involving medical physicists, that present a high degree of self-explainability/interpretability/usability and that should be more easily accepted by the clinical community. Suppose a new AI-based algorithm is used to generate a CT image: the medical physicist and the radiologist have the concrete possibility to compare the resulting output of AI application against more traditional image reconstruction methods and assess that the new way is better than the older one. Similarly, the

https://doi.org/10.1016/j.ejmp.2022.04.003
Received 17 March 2022; Accepted 9 April 2022
Available online 21 April 2022

resulting segmentation of organs and structures based on previously trained AI-based algorithms can be directly checked (and, if necessary, corrected) by radiation oncologists usually performing this task. In several situations, AI-based tools will clearly translate in the sparing of a large amount of time that doctors or other professionals may dedicate to other, less repetitive, duties: this is the case, for instance, of AI-based segmentation and radiotherapy plan optimization. The recognition that these tools work similarly or better than humans in such applications is expected to be relatively fast, right due to the definite possibility to validate and verify their performances in the clinical environment extensively. Looking at this kind of applications, we may state that we can move the elephant out of the room. This does not mean that the acceptance will be easy and rapid everywhere, primarily due to the associated changes in mentality, professional roles and responsibilities. In this context, "the room" is just our mind and the process to change the mind is never easy, especially if dealing with well-assessed practices. Within this process, the role of medical physicists is of paramount importance, principally as facilitators. Their role in imaging, radiation oncology, nuclear medicine has always been associated with innovation and implementation of new technology or new approaches [16–18]. In line with this tradition, medical physicists will facilitate the implementation of AI through rigorous validations and adaptation to local needs, mainly thanks to their specific skills and ability to understand the clinicians' language and translate, in a team effort, their requests into optimally tailored solutions [19–22].

Stop ignoring the elephant!

Coming back to the core of the Dr Kundu's message [1], let us focus on the highly topical field of radiomics. This may be considered a relevant example (among many others) of potential application of AI-based tools dealing with large amount of information extracted from images. Despite the impressively growing number of scientific publications dealing with the topic, resulting in hundreds of AI-based models for diagnosis, prognosis and outcome prediction (mostly in cancer patients), the clinical use of these results is still apparently null. In our view, this is one of the major pieces of evidence of the consequences of ignoring our elephant. Not by chance, in most of the potential applications of AI black-box radiomic scores, doctors should decide if one patient has or does not have a cancer or if they should or should not deliver therapy based on something intrinsically unexplainable. And, as we may expect, this does not happen! The sooner we recognize the issue, the sooner we find solutions. Medical physicists are part of the game; they can contribute to exploring smart ways to exploit high computational and modelling skills without forgetting the indispensable need to drive the picture toward model's explainability, interpretability and usability.

Post-hoc explainability: Is it really a solution?

Intending to make AI-based models for medicine more friendly and possibly more explainable, many researchers focus on developing methods and tools that may offer the possibility to capture the interdependence among predictors after a "black-box" model is built. These post-hoc actions are aimed to make the black box "a bit less black". They may give the operators some possibility to explore the logic behind it and, to some extent, to "score" the relative weight of different predictors or of their combinations in assessing the most relevant variables acting on single individuals. In the best situation, these attempts may offer elements of explainability from the point of view of the "technical" building of the algorithm without any possibility to deal with the true issues regarding clinical or biological explanation/interpretation [7,8,10]. This transparency is obviously of paramount importance and any effort in this direction is relevant. However, this approach continues to ignore the elephant, being intrinsically lacking from the point of view of (clinical) explainability. We can find rare exceptions in the literature; when AI tools are forced to remain potentially explainable, resulting in few-feature models [23].

Beyond post-hoc interpretation. Interpretability-driven models: First build the model, then fit it to data

The dogmatic credence that AI is "superior" in modeling (large) clinical data is unproven and basically wrong when introducing explainability and interpretability concepts. If "superior" includes the ability of models to be explainable/interpretable (and, as discussed, this should be the case!), we should admit that the "best" models are the ones based on the understanding of basic principles explaining, at least in part, the causes of what we see. The causality, proven or supposed, is a vital characteristic of usable and clinically accepted models. This is clear when few mechanisms may be summarized as those that mainly cause the effect. A quite outdated but still (to some extent) valid example is the linear-quadratic model. We have been applying it for decades in radiotherapy, and it proved robust enough to explain most of the picture in many scenarios when providing estimates of radiation-induced effects on tumors and, in part, normal tissues [24]. More recent examples include modeling tumor regression as quantified by imaging to provide good estimates of tumor control [25] and the quantitative assessing of hypoxia through functional imaging individually predicting tumor radio-resistance [26]. These are well-known examples taken from the radiotherapy field where models based on few, clear and interpretable parameters provide accurate predictions that clinicians use. Why should black-box AI data-driven models convince clinicians to replace them? Models with an intrinsic high degree of explainability align with how science progressed until our days. We may expect that the continuously improving ability to explain central basic mechanisms at the laboratory and the translational research-level would offer more and more alternatives to the black-box approach in many instances in imaging and other areas where medical physicists are deeply involved. And physicists may (and, in our opinion, should!) contribute massively.

Reducing the elephant size: Guiding out of the black-box

A major argument pushing to the use of machine learning and AI to model medical data (including medical images, laboratory data, "omics"…) is that, for the first time in the history, humans can access, pool and handle incredibly large amounts of quantitative information. A way to exploit data without forgetting the fundamental issues of explainability and interpretability is the application of machine learning and AI constrained to search the "major players". We can find examples of such an approach in the TRIPOD guidelines [27] or in the attempts to score radiomic-based models strongly weighting the model explainability [28]. Methods to reduce statistical redundancy, to select variables better explaining the "signal" and to discard features describing the "noise" (i.e. limiting overfit) can be applied in combination with machine learning and deep learning to make models more robust and generalizable as well as potentially more explainable. This approach also, at least, ensures the "control" on the effect/effect-size of treatment variables which play a key role in predictive models used in medicine. Treatment variables usually "cause" effects, they are not "casually associated" to the outcome of interest; so, we should not accept a model predicting a reduced outcome for an increase in the level of the "causing" effect [23]. Of course, this "hybrid" approach does not assure that a complete explainability and interpretability of the resulting models is achieved. Still, it looks like a good compromise in many cases. In the case not all selected best predictors are interpretable, hypotheses can be generated. They can prompt researchers to try to explain why the specific (few) variables were found to be predictive [10,29–31]. On the other hand, black-box AI-based models will never be explainable, even in the case of high performances of the resulting models.

Some concluding remarks: The elephant unveiled

AI is radically changing a large variety of activities and our societies rapidly, and none can fully predict what this will mean in the next decades. It is pretty natural that medicine will be largely influenced, maybe in a disruptive way, with potentially large benefits for the community and the individuals. This optimistic vision does not contradict our discussion: the doctors' caution in moving toward AI is related to the intrinsically ethic basement of medicine that always needs to demonstrate that "a progress is a progress" if patients benefit from it. This is a quite obvious but seldomly forgotten statement: the patient is the center both individually and at larger scales (hospital, regional, national, worldwide). Efficiency, rapidity and cost cutting (all promises of large-scale AI applications) do not translate automatically in better and sustainable care, including the reduction of inequalities in the access to adequate cures. Doctors need to trust in something that can be understood and explained without any new dogmatic absolutism of AI over human intelligence and human empathy. This claims an urgent need for a "democratic" development process for AI implementation in medicine [32]. The process will empower health practitioners against undesired effects of automated decisions, increase the trust of patients and doctors and help people make better decisions. Within this process, medical physicists have to play a relevant role in *"opening up the black box"* as Dr Kundu writes in concluding her excellent editorial.

References

[1] Kundu S. AI in medicine must be explainable. Nat Med 2021;27:1328.
[2] https://en.wikipedia.org/wiki/Shinjini_Kundu.
[3] He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25(1):30–6.
[4] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44–56.
[5] Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities. Brit J Gen Pract 2018;68(668):143–4.
[6] Stupple A, Singerman D, Celi LA. The reproducibility crisis in the age of digital medicine. npj Digit. Med 2019;2(1).
[7] Chen JH, Asch SM. Machine learning and prediction in medicine-beyond the peak of inflated expectations. N Engl J Med 2017;376(26):2507–9.
[8] EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.
[9] Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nature Med 2019;25(9):1337–40.
[10] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine (2019) Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2019;9(4):art. no. e1312.
[11] Tonekaboni S, Joshi S, McCradden M, Goldenberg MD. What clinicians want: contextualizing explainable machine learning for clinical end use. PMLR 2019:1–21.
[12] Unkelbach J, Bortfeld T, Cardenas CE, Gregoire V, Hager W, Heijmen B, et al. The role of computational methods for automating and improving clinical target volume definition. Radiother Oncol 2020;153:15–25.
[13] Wang G, Zhang Y, Ye X, Mou X. Machine learning for tomographic imaging. IOP Publishing, 2020 Bristol, UK.
[14] Wang M, Zhang Q, Lam S, Cai J, Yang R. A review on applications of deep learning algorithms in external beam radiotherapy automated treatment planning. Front Oncol 2020;19:art 580919.
[15] Pallotta S, Marrazzo L, Calusi S, Castriconi R, Fiorino C, Loi G, et al. Implementation of automatic plan optimization in Italy: status and perspectives. Phys Med 2021;92:86–94.
[16] Fiorino C, Jeraj R, Clark CH, Garibaldi C, Georg D, Muren L, et al. Grand challenges for medical physics in radiation oncology. Radiother Oncol 2020;153:7–14.
[17] Fiorino C, Guckenberger M, Schwarz M, Heide UA, Heijmen B. Technology-driven research for radiotherapy innovation. Mol Oncol 2020;14(7):1500–13.
[18] Kevill S. Physics and medicine: an historical perspective. The Lancet 2011;379:1517–24.
[19] Bosmans H, Zanca F, Gelaude F. Procurement, commissioning and QA of AI based solutions: An MPE's perspective on introducing AI in clinical practice. Phys Med 2021;83:257–63.
[20] Beckers R, Kwade Z, Zanca F. The EU medical device regulation: Implications for artificial intelligence-based medical device software in medical physics. Phys Med 2021;83:1–8.
[21] Balagurunathan Y, Mitchell R, El Naqa I. Requirements and reliability of AI in the medical context. Phys Med 2021;83:72–8.
[22] Zanca F, Hernandez-Giron I, Avanzo M, Guidi G, Crijns W, Diaz O, et al. Expanding the medical physicist curricular and professional programme to include Artificial Intelligence. Phys Med 2021;83:174–83.
[23] Carrara M, Massari E, Cicchetti A, Giandini T, Avuzzi B, Palorini F, et al. Development of a ready-to-use graphical tool based on artificial neural network classification: application for the prediction of late fecal incontinence after prostate cancer radiation therapy. Int J Radiat Oncol Biol Phys 2018;102(5):1533–42.
[24] McMahon SJ. The linear quadratic model: usage, interpretation and challenges. Phys Med Biol 2018;64:01TR01.
[25] Fiorino C, Gumina C, Passoni P, Palmisano A, Broggi S, Cattaneo GM, et al. A TCP-based early regression index predicts the pathological response in neo-adjuvant radio-chemotherapy of rectal cancer. Radiother Oncol 2018;128(3):564–8.
[26] Thorwarth D, Welz S, Mönnich D, Pfannenberg C, Nikolaou K, Reimold M, et al. Prospective evaluation of a tumor control probability model based on dynamic 18F-FMISO PET for head and neck cancer radiotherapy. J Nucl Med 2019;60(12):1698–704.
[27] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162(1):55–63.
[28] Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14(12):749–62.
[29] Astaraki M, Yang G, Zakko Y, Toma-Dasu I, Smedby Ö, Wang C. A comparative study of radiomics and deep-learning based methods for pulmonary nodule malignancy prediction in low dose CT images. Front Oncol 2021;11:art. no. 737368.
[30] Miller DD. The medical AI insurgency: what physicians must know about data to practice with intelligent machines. npj Digit Med 2019;2:62.
[31] Van Der Schaaf A, Langendijk JA, Fiorino C, Rancati T. Embracing phenomenological approaches to normal tissue complication probability modelling: a question of method. Int J Radiat Oncol Biol Phys 2015;91:468–71.
[32] Hengstler M, Enkel E, Duelli S. Applied artificial intelligence and trust: the case of autonomous vehicles and medical assistance devices. Techn Forecast Soc Changes 2016;105:105–20.

Claudio Fiorino[a,*], Tiziana Rancati[b]

[a] *Medical Physics Department, San Raffaele Scientific Institute, Milano Italy*

[b] *Prostate Cancer Program, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy*

[*] Corresponding author.

E-mail address: fiorino.claudio@hsr.it (C. Fiorino).