15 Artificial Intelligence and the Nuclear Medicine Physician: Clever Is as Clever Does

Roland Hustinx

Contents

For several years now, the role and place of artificial intelligence (A.I.) in radiology have been discussed and debated in all strata of the radiological field. From university hospitals to private centers, from large companies to countless start-ups, from scientific societies to medical associations, all are very actively and vocally involved. The U.S. Centers for Medicare and Medicaid Services' (CMS) decision in September 2020 to provide its first-ever reimbursement of a radiology A.I. algorithm is expected to open the door to broader coverage of imaging A.I. software in the clinics. The feeling in radiology is that A.I. is no longer a prospect, it is a reality. The physician's attitude has shifted from the fear that "A.I. will replace radiologists" to the belief that "radiologists who use AI will replace those who don't." A.I. has been much less present in the field of nuclear medicine (NM), which is distinct from radiology as a medical specialty in most countries. However, they share similar technologies, in particular the cross-sectional techniques used in hybrid imaging, e.g. CT and MRI. There is no reason that the advances, solutions, and new problems highlighted by A.I. in the radiological field should not be observed sooner or later in the NM field. Some of our practical specificities, such as the complication of dealing with short-lived isotopes for scheduling the clinical activity, or the complexities of individual dosimetry in treatments with radiopharmaceuticals, should, on

R. Hustinx (✉)
Division of Nuclear Medicine and Oncological Imaging, University Hospital of Liège, GIGA-CRC In Vivo Imaging, University of Liège, Liege, Belgium
e-mail: rhustinx@chuliege.be

P. Veit-Haibach, K. Herrmann (eds.), *Artificial Intelligence/Machine Learning in Nuclear Medicine and Hybrid Imaging*, https://doi.org/10.1007/978-3-031-00119-2_15

the contrary, constitute excellent fields where A.I. helps our practice. Nonetheless, it is indisputable that NM is lagging behind radiology in the clinical implementation of A.I. Whatever the reasons, increased susceptibility of the NM techniques to local methodological variables, difficulty to gather large curated datasets or perhaps smaller market less attractive for the industry, we do not seem close to seeing any reimbursement of an A.I. add-on in our field. It is only a matter of time, however, and it should give NM physicians the opportunity to better prepare and contribute more actively to shaping how A.I. will be integrated into our practice. The question is essentially twofold: what would be the role of NM physicians in a medical era where A.I. is more and more present, and what must we learn and do to shape this future.

In this chapter we shall consider successively the benefits of A.I., the threats and the obstacles that accompany its implementation, and finally the possible steps that need to be taken for a successful and mutually satisfactory embedment of A.I. in clinical nuclear medicine. These questions shall be considered looking at the three axes of involvement of A.I. in the field of NM: Physics, i.e. how A.I. will impact image acquisition and reconstruction; operational, i.e. how A.I. will optimize health care delivery through improved scheduling and overall organization; clinical which encompasses all applications aiming at improving the interpretation of the studies (not limited to the images) in terms of diagnostic accuracy, prognostic and predictive value or individual pre-treatment dosimetry.

15.1 I Am Looking Forward to More A.I. in My Practice Because…

15.1.1 The Images Will Look Prettier

In theory, we nuclear medicine physicians should benefit from the introduction of A.I. in all three fields, and the physics applications are probably the most obviously welcome. Indeed, we will be looking at images obtained with lower injected activity, i.e. lower patient's exposure [1]. Studies will be shorter to acquire, leading to improved patient's comfort and experience, fewer movement artifacts, and also increased throughput. X-ray exposure may also be reduced by using deep learning (DL) for attenuation correction, hence removing the need for low-dose, attenuation correction only, CTs [2]. A.I. has the potential to further enhance the image quality through improvements in the co-registration of the CT and SPECT/PET parts of hybrid studies. This may have major implications in particular in studies where misregistrations may have significant clinical implications. This is the case for instance when using the diagnostic CT study along with the [^{99m}Tc]MAA SPECT/CT study for determining the activity of [^{90}Y]-labeled microspheres to inject during selective intra-arterial radiation therapy. In summary, considering the images and their content as a product, we will be working with better-quality material, and nobody would argue against that.

Furthermore, improved, faster, and more robust automated AI-based segmentation algorithms will streamline the data analysis. For instance, [^{18}F]FDG PET/CT is key in the management of diffuse large B cell lymphomas (DLBCL), and the metabolic tumor volume (MTV) appears to be a metrics that further improves its prognostic value. The current consensus tends towards using a fixed maximum standardized uptake value (SUV_{max}) threshold of 4, but even when semi-automated, the process is tedious, time-consuming, and imperfectly reproducible [3, 4]. Automated algorithms based on DL have been proposed for this task [5], and in all likelihood most of us should see those as a welcome addition to our daily routine.

15.1.2 My Life Will Be Easier

The introduction of A.I. into the operation of the NM department should also benefit to the physicians, through optimization of the resources. This has been demonstrated in radiology departments [6], and it should prove even more relevant in NM, which is dealing with isotopes, including

short-lived ones. Patients scheduling, radiopharmaceutical preparation, and report generation are operational activities all susceptible to benefit from A.I., provided that the physicians, radiopharmacists, and administrative staffs strongly contribute to framing the A.I. intervention and fully stay on top of the processes. The worst-case scenario would be an A.I.-supported take-over by non-medical, bureaucratic supervisors who would consider that A.I. provides them with all the insight needed to optimally manage an NM Department, without a significant contribution from the physicians. A basic task, often overlooked, but which is responsible for a significant waste of time for the NM physician is to recover and organize previous studies, not only in NM but also in other modalities. It is often difficult to streamline a process that involves different providers, for the PACS and the different viewers that may coexist in a department. Operational A.I. would be of great value in this setting.

15.1.3 My Patients Will Be Better Off

More generally, NM physicians are used to looking at images but also at data. Radiomics and A.I. will provide more data, more reliable data, and new ways at interpreting these data. NM should therefore be a fertile ground for these developments in diagnostic and prognostic applications in general. However, we must first study the terrain before attempting to consider the practical impacts that can be expected in clinical NM. Activity profiles are very different in academic centers and public and private services. They also vary from country to country, in Europe and across the world. Some services work primarily with single-photon NM, i.e. bone scan, myocardial perfusion scintigraphy, and a range of studies performed less frequently such as kidney, thyroid, or parathyroid scans. These studies, when added together, constitute a significant contribution to the production of these services. The relative contribution of hybrid imaging (SPECT/CT) also varies considerably from center to center. In yet other departments, most of the activity relates to PET/CT, and some regularly perform a large number of non-FDG studies, such as radiolabeled PSMA ligands. In addition, theranostic approaches, with the accompanying treatment procedures, also occupy very different places in NM centers. Therefore, it is clear that considering the potential impact of A.I. in the field of NM involves first trying to understand the major trends in the future development of the specialty itself. A systematic review published in 2019 showed a strong imbalance in A.I. applications towards oncology, which accounted for 86% of all publications in A.I. and radiomics fields [7]. Hence, one may infer that those centers where oncology, and more specifically high-end, tertiary or quaternary-care oncology, is more prevalent, will experience the most immediate impact of A.I. on their clinical practice. Neurology and cardiology are probably the next in line in terms of clinical implementation. From the physician's perspective, the initial steps in this clinical implementation process should be quite exciting. We can expect to benefit from a growing number of A.I. toolkits designed to perform dedicated and highly focused tasks, such as characterizing lung nodules using [^{18}F]FDG PET and CT, or recognizing normal patterns, e.g. non-pathological studies in whole-body bone scans with [^{99m}Tc]-labeled diphosphonates. Such tasks should prove to be of great benefit to the specialty, and our patients, by improving the quality and reliability of the diagnostic information contained in our reports. We would always maintain a holistic, human-centered approach to the NM imaging field, as we would use these A.I. tools to merely complement an otherwise unchanged process of interpreting images and quantitative data that supports them. Personalized dosimetry may also be helped by A.I. and thus gain further acceptance in the clinical field. For instance, similar to diagnostic studies, A.I. may lead to shorter acquisition times for the [^{177}Lu] SPECT studies or better model and predict voxel-wise dosimetry measurements. Again, the final decision, i.e. should we treat the patient and if yes, the activity to be administered, would remain in the physician's hands, albeit better armed for making those decisions.

With all of these largely positive elements, the transition to AI-augmented nuclear medicine should be smooth and easy. All we have to do is learn how to use the new tools first and then how extensively to trust them. Just as we use quantitative algorithms that compare individual studies to population-based normality, like the Cedar-Sinai program in MPI or Parametric Statistical Mapping (SPM) in FDG brain PET studies, and many more. These are useful tools, fully integrated into the clinics, but the conclusions of which do not replace those of the NM physician. Obviously, however, this is not the full story. Indeed A.I. undoubtedly contains threats to the practice of nuclear medicine as we know it, and as some us might want to keep it. And other obstacles exist in the way of a smooth implementation of A.I. in clinical NM.

15.2 I Am Wary of More A.I. Because…

15.2.1 I Don't Understand It

This represents perhaps the greatest obstacle on A.I.'s path towards clinical nuclear medicine. As stated previously, we as NM physicians are used to dealing with data, numbers, values, quantitative measurements in addition to looking at images. We understand the relationship between these numbers and results, and the physiological, biological, or biochemical processes that underlie them. We easily translate time/activity curves into glomerular filtration rate. We understand how to translate counts/pixel into the SUV, as a semi-quantitative measurement of the glucose metabolism. We also understand and know very well all the factors that affect the variability of the SUV. We also know that we could, if we wanted to, obtain absolute measurements such as the glucose metabolic rate in mmol/min/g. tissue. Every nuclear medicine physician knows the difference between filtered back-projection and iterative reconstruction. We have been trained to master the basics of physics and instrumentation, and we are able to speak or at least listen to our fellow physicists and engineers. However, our training in computational science and our understanding of probabilistic learning is quite limited. For many of us, the leap to radiomics is reasonably doable, because they are quantitative features that answer formulas, and for which we can assess confounders. Basically, the good old SUV is nothing more than a basic radiomic function. The more advanced features remain very similar whether they represent a measure of signal heterogeneity, shape or intensity, e.g. the biological phenomenon responsible for the accumulation or distribution of the tracer. The leap to A.I. is much more difficult, because our scientific background has not prepared us for it. We do not have the mental tools to fully understand the basics of a U-Net architecture. Without even considering DL, the more basic learning machine algorithms, such as the random forests and support vector machine, are not entirely part of our natural domain of competence. Furthermore, the relationship between the images, the quantitative features abstracted from the images and the biology, is lost after going through the DL process. Moreover, with A.I. in medicine, high performance is often associated with high opacity. Hence the call for explainable and interpretable A.I. Some authors have gone further in distinguishing explainability and causability [8]. The former "highlights decision-relevant parts of the used representations of the algorithms and active parts in the algorithmic model, that either contribute to the model accuracy on the training set, or to a specific prediction for one particular observation." The later refers to "the extent to which an explanation of a statement to a human expert achieves a specified level of *causal understanding* with effectiveness, efficiency and satisfaction in a specified context of use." In other words, an algorithm is explainable if we understand the effect of variables on all the moving parts that constitute the algorithm, and it fits the causability criterion if the end result, i.e. the conclusion at the end of the computation, is efficiently and transparently actionable.

15.2.2 I Don't Trust It

Obviously, it is difficult to trust processes that are poorly understood, which is why explainability and causability are prerequisites for trust. Beyond that, A.I. is not free of risk, in particular it can generate errors. For example, image reconstruction with DL can lead to artifacts and alterations that could have clinical impact [9]. Machine learning algorithms, even the smartest, can be fooled by minute alterations to the input data and completely mishandle the data, in a way that humans are not subject to [10]. This is the so-called "adversarial machine learning" well known in the A.I. community, and the concept has been extended to the field of radiomics [11]. This raises the specter of an initially effective and fully validated A.I. algorithm turning into a mill generating mislead interpretations and erroneous decisions. The validation process itself needs to be validated. The medical literature is not devoid of papers that, although peer-reviewed in a seemingly appropriate fashion, are methodologically impaired in a severe way. Many questions arise concerning the statistical methods for assessing the performance of an algorithm. Most articles in NM use the area under the receiver operating characteristic curve (AUC ROC) as the main metric for assessing the performance of the model when the outcome is binary, i.e., recurrence/no recurrence, malignant/not malignant, etc. Yet in presence of unbalanced data, the AUC artificially inflates the performance of the model [12]. There is a need for at the very least using the most appropriate test, e.g. AUC and F-score, depending on the sample distribution and hypothesis, and also probably to develop more specific tests [13].

Further improving and perfecting the A.I. should be accompanied by further safeguards. Current typical A.I. models are essentially static, in that they have been trained using samples corresponding to a population that was fully validated at the time the model was built. They are efficient in test sets that correspond to their training sets. Those static algorithms may be subject to concept drift, which means that even though a task was at first efficiently and reliably fulfilled, it is no longer the case when the patient population evolves or when the technique changes. So ideally, the algorithms should not stop learning, i.e. they should adapt along with modifications introduced in the sets of data to analyze. This is the continuous learning or continual A.I. [14]. The algorithm learns to learn, incrementally adapts to new characteristics found in the input data, constantly updating its feature selection to better fit its changing environment. Intuitively we may realize the advantages of such process, but we also realize that it should be associated with a constant "revalidation process." Indeed, the catastrophic inference or forgetting may occur when extreme outliers wreak havoc into an autonomously relearning algorithm. To put it simply, even fully validated and trustworthy A.I. algorithms at the time of marketing and clinical implementation need to continuously go through extremely stringent quality controls.

15.2.3 I Don't Want It

The ultimate, and most compelling, question is "where does the physician fit in this puzzle?" Say we end up with a multitude of A.I. algorithms dedicated to a multitude of specific tasks, possibly running in parallel and selected depending on the patient's medical profile and issue at hand. Say those algorithms are constantly learning, and one way or another, the process is safeguarded by multiple checkpoints. Once we get there, the role of the physician could go either way: The physicians remain in charge of the patient's care, responsible before the law, they keep receiving the medical fees, and thus decide when and how to use the A.I. tools. Or the physicians do not have the knowledge and expertise to correct the A.I. tools when they are wrong; they do not even know when an A.I. tool is wrong, and they are surrounded by so many effective A.I. tools that the gestalt, which was the heart of the medical profession, is no more than the vestige of a bygone era, in so much so that the physicians no longer enjoy the confidence of the public and health care providers. The debate remains very vivid in the radiology community. The prophecy

G. Hinton playfully made in 2016 ("People should stop training radiologists now. It's just completely obvious within five years that DL is going to do better than radiologists") has not been verified yet, but the question remains circulated in the decision circles. The Dutch Finance Minister Wopke Hoekstra very recently commented that "The work of the radiologist to a significant extent has become redundant, because … a machine can read the images better than humans who studied 10 years for it" [15]. The answers coming from medical and scientific organizations are only half-convincing. They argue that as the medical demand is increasing, A.I. will take care of the automated, time-consuming tasks, always in support of the physicians, whose number will remain stable, hence improving the cost/effectiveness ratio of the radiological profession. They add that "AI will still make mistakes, which can be easily corrected by a human, by a radiologist. But will not be possible for AI to correct itself" [15], which as we have seen represents more wishful thinking than hard truth. Furthermore, considering the balance "who corrects who," past experience with computer-assisted diagnosis is not uniformly encouraging as, in some instances, radiologists tend to ignore or overturn the computer prompts, even when they are correct [16]. Needless to say, implementation of A.I. in the clinics has massive implications in terms of legal responsibilities, but this topic would deserve a full chapter.

15.3 How to Proceed? Let's Be Practical!

Radiology is ahead of nuclear medicine, and seems caught in a circular argument: A.I. is there to stay, it's going to be faster, more powerful, and more reliable for organizing the departments and providing the clinicians with the most relevant information, yet radiologists need to remain totally in charge and in full control.

The key issues are probably the validation of the A.I. algorithm and its endpoint. A typical approach is to compare the A.I. with the human truth. A good example is provided by Sibille et al. who identified, located, and segmented over 12,000 regions in 629 FDG PET/CT studies performed in lymphoma and NSCLC patients [5]. A DL algorithm using both the PET and the CT data performed very well for these tasks, with 87.1% sensitivity and 99% specificity in classifying the lung cancer patients, and 88.6% localization accuracy in the same population. Similar results were obtained in the lymphoma patients. In this case, the network is trained to do as well as the physician. It does not reach this level of performance, but close enough, and is thus proposed as an adjunct to the physician's interpretation. In this case, we do not know the ground truth, we do not know who is right in the discrepant cases (human "gold standard" or DL?), but it does not matter, as the product is designed to help the physician accomplishing his task, including the potential flaws. This is a very marketable product, because it does not change the paradigm, the physician remains in charge, and the product being a tool that automates and accelerates a process. It has been trained to replicate the human's process, and it is designed to be checked by humans.

Following this approach does not fully take advantage of the capacities of A.I. Zhao et al. recently went further with their report on DL for diagnosing metastatic involvement on bone scintigraphy [17]. They studied over 12.000 cases, and the endpoint was clear-cut, i.e. the presence or absence of bone metastases in the scintigraphy. They showed an overall accuracy of 93.4%, with 92.6% sensitivity and 93.9% specificity and an AUC of 0.964, consistent across cancer types. This compared favorably with the performances of experimented NM physicians, as in 13/200 cases read in parallel, A.I. was correct and all three physicians were wrong, compared to only 6 cases where it was the reverse. And this was obtained at lightning speed, as only 11 seconds were needed for interpreting 400 cases, which is…fast! As a comparison, it took an average of 136 minutes for the NM physicians to read those 400 studies, e.g. almost 3 studies per minute, which for a human being, is also quite fast. This paper is a good case study. Published in a prestigious journal, the conclusion is unequivocal: A.I.

is faster, better, and cheaper than the physicians. Case closed. In this model, there is no need for a physician in control, no A.I. at the service of the physician, and no A.I. as a complement or support to the physician. A.I. wins, period. Yet in order to go further and implement such algorithm in the clinic, one must first answer a few questions. The study deals with planar scintigraphy, although SPECT is recommended and routinely performed. That is relevant because the benefit of A.I. was primarily in terms of sensitivity. Also, adding the CT further improves the diagnostic accuracy. The ground truth is also debatable, as explained in the methods. And finally, the algorithm is the perfect example of a black box. Hence, this tremendous amount of work (over 12.000 studies!) published in a high-level journal, provides very little chance of effective clinical translation, if NM physicians are asked to give their opinion. The imaging technique is not up to date, the gold standard is weak, the method is questionable, and the algorithm is opaque. Similarly to some extent, major critiques were addressed after the publication of a paper reporting on a DL algorithm outperforming radiologists for interpreting mammographies, even though this study was methodologically very solid [18, 19]. One may wonder whether A.I., to be accepted, must be clamped and its power limited.

In order to get out of this labyrinth and come to the situation where not only nuclear medicine physicians coexist with A.I. but patients also truly benefit from this development, a multistep approach is required. First, physicians must identify unmet clinical needs, taking into account the bigger picture. This means identifying the weak points of our techniques, in terms of accuracy or reproducibility, in diseases and clinical situations where it makes a difference for patients. [^{18}F] FDG-PET/CT is quite effective in identifying residual disease at the end of treatment for diffuse large B-cell lymphoma. The advantage of developing A.I. for this task would be marginal at best, and difficult to establish. The impact would be quite different were it to predicting or assessing early response to immunotherapies, which can be very effective but in a limited number of patients and with significant costs, both monetary and in terms of morbidity. Theranostics is a major field for the development of A.I. in nuclear medicine, to help the physicians in identifying those who would benefit from the treatment based upon the diagnostic companion study, tailor the treatment through fast personalized dosimetry, and finally reliably and rapidly assess treatment success, or failure. Second, we need to acquire the minimal knowledge necessary to get on speaking terms with those who will actually develop and build A.I. This goes through changing how the research teams are organized, developing strong collaborations outside the faculty of medicine, and probably partnering with the industry. This also implies revamping the education and training of residents to account for this evolution. We have to get better in statistics and computational sciences. Third, we need to build multicenter networks. It is very unlikely that single-center protocols will manage to gather the amount and diversity of data necessary to develop A.I. algorithms directly applicable to the routine clinical practice. We need to account for the diversity in the hardware performances, acquisition and reconstruction algorithms, and population types. And finally, we need to set the highest standards for validation, not only regarding the methodology surrounding the development and testing of the A.I. model but also the clinical relevance of the question being solved and the clinical appropriateness of the population sample being investigated.

If we can fulfill these criteria, i.e. if we identify the need, comprehend the methods, and put ourselves in a situation such as to produce reliable and reproducible results, then and only then will we be fully prepared for the next phase, i.e. enthusiastically promoting and advocating the A.I.-augmented nuclear medicine to the clinical world.

Acknowledgements The author wishes to thank Nadia Withofs, MD, PhD for fruitful discussions.

References

1. Schwyzer M, Ferraro DA, Muehlematter UJ, Curioni-Fontecedro A, Huellner MW, von Schulthess GK, et al. Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks—initial results. Lung Cancer. 2018;126:170–3. https://doi.org/10.1016/j.lungcan.2018.11.001.
2. Shiri I, Ghafarian P, Geramifar P, Leung KH, Ghelichoghli M, Oveisi M, et al. Direct attenuation correction of brain PET images using only emission data via a deep convolutional encoder-decoder (Deep-DAC). Eur Radiol. 2019;29:6867–79. https://doi.org/10.1007/s00330-019-06229-1.
3. Burggraaff CN, Rahman F, Kassner I, Pieplenbosch S, Barrington SF, Jauw YWS, et al. Optimizing workflows for fast and reliable metabolic tumor volume measurements in diffuse large B cell lymphoma. Mol Imaging Biol. 2020;22:1102–10. https://doi.org/10.1007/s11307-020-01474-z.
4. Barrington SF, Zwezerijnen BG, de Vet HC, Heymans MW, Mikhaeel NG, Burggraaff CN, et al. Automated segmentation of baseline metabolic total tumor burden in diffuse large B-cell lymphoma: which method is most successful? J Nucl Med. 2021;62(3):332–7. https://doi.org/10.2967/jnumed.119.238923.
5. Sibille L, Seifert R, Avramovic N, Vehren T, Spottiswoode B, Zuehlsdorff S, et al. (18)F-FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks. Radiology. 2020;294:445–52. https://doi.org/10.1148/radiol.2019191114.
6. Curtis C, Liu C, Bollerman TJ, Pianykh OS. Machine learning for predicting patient wait times and appointment delays. J Am Coll Radiol. 2018;15:1310–6. https://doi.org/10.1016/j.jacr.2017.08.021.
7. Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. Eur J Nucl Med Mol Imaging. 2019;46:2656–72. https://doi.org/10.1007/s00259-019-04372-x.
8. Holzinger A, Langs G, Denk H, Zatloukal K, Muller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov. 2019;9:e1312. https://doi.org/10.1002/widm.1312.
9. Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. Proc Natl Acad Sci U S A. 2020;117(48):30088–95. https://doi.org/10.1073/pnas.1907377117.
10. Zhou Z, Firestone C. Humans can decipher adversarial images. Nat Commun. 2019;10:1334. https://doi.org/10.1038/s41467-019-08931-6.
11. Barucci A, Neri E. Adversarial radiomics: the rising of potential risks in medical imaging from adversarial learning. Eur J Nucl Med Mol Imaging. 2020;13:2941–43. https://doi.org/10.1007/s00259-020-04879-8.
12. Cook J, Ramadas V. When to consult precision-recall curves. The Stata Journal. 2020;20:131–48. https://doi.org/10.1177/1536867x20909693.
13. Flach P. Performance evaluation in machine learning: the good, the bad, the ugly, and the way forward. In: The thirty-third AAAI conference on artificial intelligence (AAAI-19). 2019.
14. Pianykh OS, Langs G, Dewey M, Enzmann DR, Herold CJ, Schoenberg SO, et al. Continuous learning AI in radiology: implementation principles and early applications. Radiology. 2020;297:6–14. https://doi.org/10.1148/radiol.2020200038.
15. Turner J, Ward P. Dutch debate intensifies over future shape of AI. 2020. https://www.auntminnieeurope-com/indexaspx?sec=sup&sub=aic&pag=dis&ItemID=619384.
16. Nishikawa RM, Schmidt RA, Linver MN, Edwards AV, Papaioannou J, Stull MA. Clinically missed cancer: how effectively can radiologists use computer-aided detection? AJR Am J Roentgenol. 2012;198:708–16. https://doi.org/10.2214/AJR.11.6423.
17. Zhao Z, Pi Y, Jiang L, Xiang Y, Wei J, Yang P, et al. Deep neural network based artificial intelligence assisted diagnosis of bone scintigraphy for cancer bone metastasis. Sci Rep. 2020;10:17046. https://doi.org/10.1038/s41598-020-74135-4.
18. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577:89–94. https://doi.org/10.1038/s41586-019-1799-6.
19. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Shraddha T, Kusko R, et al. Transparency and reproducibility in artificial intelligence. Nature. 2020;586:E14–6. https://doi.org/10.1038/s41586-020-2766-y.