



Physician centred imaging interpretation is dying out — why should I be a nuclear medicine physician?

Roland Hustinx^{1,2}

Received: 2 May 2019 / Accepted: 23 May 2019 / Published online: 7 June 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Radiomics, machine learning, and, more generally, artificial intelligence (AI) provide unique tools to improve the performances of nuclear medicine in all aspects. They may help rationalise the operational organisation of imaging departments, optimise resource allocations, and improve image quality while decreasing radiation exposure and maintaining qualitative accuracy. There is already convincing data that show AI detection, and interpretation algorithms can perform with equal or higher diagnostic accuracy in various specific indications than experts in the field. Preliminary data strongly suggest that AI will be able to process imaging data and information well beyond what is visible to the human eye, and it will be able to integrate features to provide signatures that may further drive personalised medicine. As exciting as these prospects are, they currently remain essentially projects with a long way to go before full validation and routine clinical implementation. AI uses a language that is totally unfamiliar to nuclear medicine physicians, who have not been trained to manage the highly complex concepts that rely primarily on mathematics, computer sciences, and engineering. Nuclear medicine physicians are mostly familiar with biology, pharmacology, and physics, yet, considering the disruptive nature of AI in medicine, we need to start acquiring the knowledge that will keep us in the position of being actors and not merely witnesses of the wonders developed by other stakeholders in front of our incredulous eyes. This will allow us to remain a useful and valid interface between the image, the data, and the patients and free us to pursue other, one might say nobler tasks, such as treating, caring and communicating with our patients or conducting research and development.

Keywords Artificial intelligence · Radiomics · Nuclear medicine · Molecular imaging

Introduction

“People should stop training radiologists now. It’s just completely obvious within five years that deep learning is going to do better than radiologists. It might take ten years, but we’ve got plenty of radiologists already. I said this at a hospital, and it didn’t go down too well” (https://www.reddit.com/r/medicalscool/comments/6nlgta/ai_expert_stop_

[training_radiologists_your_thoughts/](#)). These words, uttered by Geoffrey Hinton during a meeting in November 2016, sent shockwaves through the medical imaging community. Indeed, Hinton is not your usual pundit. He is a renowned cognitive psychologist and computer scientist and considered the “godfather of neural networks”. Ever since this statement, those involved in the field of artificial intelligence (AI) have gone to great length to soften the stance when debating its impact on the practice of radiology. Nowadays, the motto is no longer “AI will replace radiologists”, rather, “radiologists using AI will replace those who don’t”, which is much more politically correct and might likely lead to better business opportunities.

So, why should I be a nuclear medicine physician (or a radiologist)? Let us first consider what exactly is the job of a nuclear medicine physician, and, second, let us try to envision how, for better or worse, AI will affect this job. There are no easy answers to these questions, since, to quote Hinton again, “It is very hard to predict beyond five years in this area and

This article is part of the Topical Collection on Advanced Image Analyses (Radiomics and Artificial Intelligence).

✉ Roland Hustinx
rhustinx@chuliege.be

¹ Division of Nuclear Medicine and Oncological Imaging, University Hospital of Liège, Liège, Belgium

² GIGA-CRC in vivo Imaging, University of Liège, Sart Tilman, B35, 4000 Liège, Belgium

things always turn out differently to what you expect” (<https://www.telegraph.co.uk/technology/2017/08/26/godfather-ai-making-machines-clever-whether-robots-really-will/>).

According to the definition provided by the European Union of Medical Specialists, “Nuclear Medicine is a branch of medicine that uses unsealed radioactive substances for diagnosis and therapy.... The procedures within the scope of this definition include in vivo imaging with radiopharmaceuticals, correlative/multimodality imaging, radionuclide-guided surgery, dosimetry, therapy with radioactive substances, techniques related to nuclear physics in medicine, as well as the medical applications of radiobiology, in vitro procedures and radiation protection” (<http://uems.eanm.org/index.php?id=39>). Our specialty is wonderfully diverse, structured by and relying primarily on radiopharmaceuticals, hence on biology, and cannot be reduced to image viewing or image interpretation. Considering the diagnostic aspects of nuclear medicine, we have to deal with functional images that could be dynamic, static, whole-body, tomographic, and often a combination of the above. Increasingly, these functional images are associated with radiological images, usually computed tomography (CT), but not exclusively. Thyroid ultrasound, for example, is widely performed by nuclear medicine physicians, and magnetic resonance imaging (MRI) has entered our field through positron emission tomography (PET) / MRI, albeit essentially in a research setting. Whether these radiological images are fully interpreted or merely serve as adjuncts to the functional images, nuclear medicine physicians have gained increased familiarity with this aspect of medical imaging. For all intents and purposes, the effect of AI is likely to be very similar in both nuclear medicine and radiology with regard to their diagnostic applications. Beyond the images, however, we also extract functional information expressed in clinically meaningful values, such as the glomerular filtration rate, the left ventricular ejection fraction, and the gastric emptying rate. The therapeutic applications of nuclear medicine are numerically smaller than their diagnostic counterpart, but this field has seen recent clinical and scientific developments and constitutes an important component of nuclear medicine’s future. Indeed, the demand for radioisotopes is increasing 5% annually (<http://www.world-nuclear.org/information-library/non-power-nuclear-applications/radioisotopes-research/radioisotopes-in-medicine.aspx>).

AI will be helpful even before the nuclear medicine physician starts looking at an image

Although data remain very scarce, AI could help streamline the workflow and, possibly, increase patients’ safety. Analysing patients’ electronic medical records may, for instance, help predict no-shows, and thus better target patients,

who should be more actively reminded of their appointment [1, 2]. This could be of particular interest in nuclear medicine, where radio-isotopes are a highly valuable resource with, by definition, a limited shelf life. Similarly, machine learning and text-mining could help screen patients at risk [3].

It appears increasingly likely that AI, and in particular deep learning, will become routinely implemented in PET image generation. One of the aims would be to decrease patients’ exposure by lowering the injected activity while maintaining image quality and quantitative accuracy [4] or using pseudo-CT data for attenuation correction [5]. The potential is real for AI methods to contribute to better organisation of nuclear medicine departments and optimise resources. As always with AI at this stage, the operative word is “potential”, because only feasibility studies are currently available.

AI will improve the diagnostic accuracy of our tests

Fluorodeoxyglucose (FDG) PET/CT is a success story for nuclear medicine, and it is widely integrated in clinical algorithms, especially in oncology, where it is strongly supported by high-level evidence [6]. Of course, even in the diseases where it is considered a cornerstone of the diagnostic work up, there is room for improvement in terms of diagnostic accuracy. For instance, FDG PET/CT is widely recommended for the preoperative staging of non-small cell lung cancer, yet the overall sensitivity and specificity for mediastinal nodal assessment are 77.4–81.3% and 90.1–79.4%, respectively, depending on the criteria defining positivity [7]. For assessing nodal status in head and neck cancer 6 months after treatment, the sensitivity is 85% (76–91%), and specificity 93% (89–96%), and with the prevalence set at 10%, the negative predictive value is almost perfect (98%), but the positive predictive value is 58% [8]. The point being made here is not to undercut the clinical value of FDG PET/CT, but to illustrate that even in clinical situations where the technique is very well implanted with a clinical evaluation based on abundant literature and clinical experience, it still might be possible to further improve the accuracy. Similarly, myocardial perfusion imaging with single photon emission computed tomography is widely used for the diagnosis of obstructive coronary artery disease, yet, with reported sensitivity and specificity of 88% and 61%, respectively, it is not as if the performances could not be improved [9]. Technological hardware developments, such as CT-based attenuation correction or Cadmium Zinc Telluride detectors may help, but not to the point where further progress should not be sought [10, 11].

Diagnostic accuracy could be improved in essentially two ways: increasing the detection of anomalies and improving the characterisation of these anomalies. In the fields of radiology and pathology, there is already ample evidence supporting the

effectiveness of AI in these tasks [12–14]. In nuclear medicine, results are scarce and remain highly preliminary, but they are encouraging, whether it concerns PET/CT with FDG [15], other tracers [16, 17], or single-photon emission computed tomography (SPECT) [18, 19]. Cognitive decline related to Alzheimer's disease is associated with patterns of impaired brain glucose metabolism, which are, in the clinical setting, visually assessed, sometime with the help of statistical mapping or other quantitative analyses. Choi et al. recently proposed to develop an automatic image interpretation system based on a deep convolutional neural network (CNN) [17]. Such method does not require manually defining image feature extraction, and used images that were not further processed in terms of spatial normalisation or segmentation for instance. Combining the information extracted from FDG and Florbetavir studies, an amyloid tracer, conversion to Alzheimer's disease from minimal cognitive impairment was predicted with an accuracy of 84.2%, which compared favourably to other quantitative methods currently available. This is an example of a diagnosis that is automatically provided by an algorithm, with very minimal involvement by the physician.

In FDG PET, the area under the curve (AUC) for characterising solitary pulmonary nodules has been reported in the range 0.87–0.94 [20, 21]. Using deep learning, the AUC was 0.989 for distinguishing lung cancer patients from normal FDG PET/CT studies [22]. The question was asked in a binary fashion, i.e. cancer/non-cancer. The AUC remained extremely high (0.97) in images reconstructed using 3.3% of the counts, which in the clinical setting could translate into very low injected activity. As with most of these studies, the results must be taken with a grain of salt, because only 100 patients, including 50 with lung cancers (23 with stage IV disease), were studied.

AI will improve the diagnostic accuracy of our readers

The diagnostic accuracy reported in the literature is obtained through prospective or retrospective studies. In the latter case, images are often re-examined in prospective fashion. Those studies do not faithfully reflect the diagnostic accuracy as observed in actual clinical practice, where accuracy may be lowered by degradations in image quality or in image interpretation. In a recent survey, referring clinicians perceived a misinterpretation, most often overinterpretations, of the findings by the reporting physician in 5–20% of oncological FDG PET/CT studies [23]. Physicians, including nuclear medicine physicians, are not infallible, they do make errors. An error can be defined as a “mistaken decision” or, more precisely, “a commission or an omission with potentially negative consequences for the patient that would have been judged wrong by skilled and knowledgeable peers at the time it occurred, independent of whether there were any negative consequences”

[24, 25]. The error rate in radiology and nuclear medicine is obviously difficult to ascertain, but it has been reported that 4% of routine radiologic interpretations contain errors [25]. The most common errors are perceptual and interpretative: anomalies are not recognised as such or are mistakenly interpreted. Sources and causes of errors are multiple and have been thoroughly studied [26]. Overburden of work leading to fatigue and lack of attention is a leading source of errors, but other, more trivial, factors may also contribute. For instance, interruption of a radiology resident's work by phone calls tends to decrease their diagnostic accuracy [27]. Computer-aided detection or diagnosis, which is entirely different from AI, has been proposed as an adjunct with recall rates up to 12% in the context of mammography screening [28]. Interestingly, radiologists in that study ignored the majority of the correct computer prompts. Peer review is another effective way to improve diagnostic accuracy, but it comes with a cost that is often not sustainable [29, 30]. In radiology, i.e., radiographs, CT, and MRI, double reading leads to highly variable discrepancy rates from 0.4% to 22% [31]. With FDG PET/CT, 13% of the studies reviewed by a sub specialist as a second opinion in an outside hospital saw their diagnoses correctly modified in the majority of cases [32]. Even when applied in a rigorous and appropriate fashion, the current methods may show major limitations. A recent study showed that in more than half of the cases, experienced readers selected different, yet valid, target lesions on baseline CT. This leads to very poor agreement for classifying disease evolution according to RECIST 1.1 criteria [33].

Obviously, computers are not bothered by phone calls, and deep learning algorithms are unaffected by fatigue or mood swings. Hence, their diagnostic accuracy, as reported in the literature, should be achievable in clinical practice, provided the patient populations are similar. No definite data are available regarding this aspect, but preliminary attempts at analysing myocardial SPECT studies and generating structured reports using AI are quite encouraging [34].

AI will enhance the clinical impact of our techniques

Imaging studies are conducted for a specific purpose in a context unique to the patient being investigated. Limiting the scope of AI to image acquisition and interpretation would artificially limit its impact on patient care, since it surely could extend well beyond. Considering the patient as a whole and combining clinical data, medical history, and radiomic features collected from various imaging studies along with biological, and possibly genomic, transcriptomic, and metabolomic through deep learning could lead to unique signatures that are eminently desirable in the perspective of precision medicine. We are not there yet, but the concepts are in

place, and the methodology is being defined to test those combinations in the clinical field [35, 36]. Finally, as mentioned earlier, therapeutic nuclear medicine is clearly progressing, and customised dosimetry is increasingly considered as a key to improving the outcome. Yet again, AI techniques are contributing to improving the process albeit with limited but encouraging results [37].

Why should I be a nuclear medicine physician in this context?

Basically, we are being told that AI will take care of organising our departments and scheduling our patients, define the most appropriate acquisition and reconstruction parameters, interpret the images and their quantitative substrate without being interfered with by non-medical issues, and put everything in the appropriate context for optimal decision-making and management. How the physician fits into this picture is not completely apparent, but there are emerging answers.

The first is the timeline and the validation of the processes. We currently see many promises and unlimited perspectives but very few hard facts. In this article, the term AI has been used in a generic fashion, since it refers to all the processes by which the computer mimics “cognitive” functions that humans associate with other human minds. As discussed elsewhere, machine learning is one type of AI, and it includes many different methods such as random forests, support vector machines, Bayesian networks, neural networks, etc. (see also the article of Visvikis et al. in this supplement). Deep learning is an advanced form of machine learning with the capability of analysing the entire image without the need for manually or semi-automatically identifying and segmenting lesions or areas of interest for features extraction. However, it requires very large datasets for testing and validation. Most studies to date have tested one particular method to answer a very specific question. For instance, Kirienko et al. used linear discriminant analysis (LDA) to solve the following problem: “Is this hypermetabolic lung lesion a primary tumour or metastasis?” The AUC and accuracy in the validation set were 0.91 and 81.7%, respectively, which is quite impressive [38]. Hsu et al. used random forest analysis to solve the following problem: “Considering the areas of high FDG uptake on whole-body studies, are these normal structures or tumours?” [39]. Normal tissues were recognized with a 90% accuracy, which is equally impressive. However, there are very few side-by-side comparisons of the various ML approaches such as that performed by Deist et al. [40]. Considering the multiple indications of imaging tests, we would need to systematically and thoroughly evaluate several algorithms for each indication in order to identify the most appropriate in each case. It is unlikely that one size would fit all, so packages will have to be developed following extensive validation. One must also consider the

additional findings frequently observed in imaging studies and reported by the nuclear medicine physicians. Such findings might be irrelevant with respect to the precise clinical indication of the test, but they could be very relevant to the patient’s management, and they should, therefore, be recognised. Let us consider a frequent indication, e.g. characterisation of a solitary pulmonary nodule with FDG PET/CT. We would use an algorithm designed to solve the problem “is this lung nodule benign or malignant?”. If the patient also presents with a focal colonic uptake that is left aside and not appropriately reported (adenoma? Early-stage primary?), the end-result of the PET study should be regarded as a failure, even though the problem was solved with a near-perfect accuracy.

Although it is difficult to know what the end of the road will look like, the most likely path would be for the physician to progressively alter the way he/she performs and interprets imaging and nuclear medicine studies. This gradual process implies integrating AI toolboxes the way other computer-driven changes have been integrated in the past. Considering FDG PET in solitary pulmonary nodules, we started looking at non-attenuation corrected images in essentially a visual, binary fashion. We then evaluated whether semi-quantitative measurements such as the standardized uptake value (SUV) could help, and, finally, we automated segmentation tools of the CT image at our disposal for precise nodule size measurement, which we can combine with the visual assessment of the metabolic activity using a score relative to the background. If we follow the recommendations of the British Thoracic Society, we can easily add these findings to other relevant clinical and imaging data through a predictive model and provide clinicians with a likelihood estimate of malignancy [41]. This approach is the ultimate development of the conventional image analysis and diagnostic algorithm: first, identify an anomaly; second, characterise it using visual metrics (size, activity); third, combine these simple metrics with other known predictive factors (age, gender, location, smoking history, speculation, etc.) using non-learning predictive models, in this case the Brock and the Herder models [42, 43]; fourth, draw a conclusion. Most often, we do not go that far. We just mentally compute probabilities according to the imaging characteristics that we are able to recognise and the clinical information of which we are aware. Both are highly variable. The former, the imaging characteristics, currently depends on features recognised by the human eye, and performances do vary greatly from one observer to another according to the level of experience and expertise, and they also vary within a single observer depending on their level of attention. The latter, the clinical information, depends on the quality and exhaustiveness of the information provided by the referring physician, the quality and quantity of information that could be obtained from the patient, and the access to the patient medical record including previous studies. Surely, as physicians primarily concerned with the well-being of our patients, we would welcome anything that would reliably

streamline the processing of the information, both clinical as well as that obtained in the imaging study. Of course, as physicians ultimately responsible for the information provided to the clinicians, we also want to exercise our best judgment when writing the report. From that perspective, the short-term future could not be brighter for nuclear medicine physicians/radiologists. Many additional tools will be added to our already rich armamentarium, tools that will improve image quality and reduce patient exposure. We can envision an AI-augmented clinical practice of nuclear medicine. The interpretation process will be systematised so that error rates will be reduced. Additional and original information will be extracted from diagnostic studies and also help individualise the treatments with radiopharmaceuticals. Physicians will remain in charge, but with a reduced workload and thus more time to devote to human tasks, i.e., prospective thinking especially for further research and development and privileged contacts with clinicians and patients. The benefit of the later, i.e. more time with patients is self-explanatory, but the importance of the former, i.e. research should not be overlooked. Indeed, as defined earlier, nuclear medicine is based upon radiopharmaceuticals. Progress in nuclear medicine does not only stem from enhanced spatial resolution, shorter examination time, lower radiation dose, improved quality and robustness of the interpretation, it is also, and more importantly so, derived from new radioligands targeting the most relevant biological phenomena. Increased time and resources devoted to interdisciplinary work aimed at bridging basic sciences such as molecular biology, genetics, etc. to unmet clinical needs through the development and validation of new radiotracers and therapeutic agents should be extremely welcome. This could give us a unique opportunity to further develop our techniques and establish their clinical effectiveness. Of course, this will be possible only if we remain ourselves, as medical specialists, relevant players in the field.

There are still issues to be resolved. One is the realm of the physician oversight. The decision comes with the responsibility, and the physician will always be able to overrule the results provided by an AI algorithm, just as he/she frequently discards CAD prompts even when they are correct. In other words, one will have to figure out the tipping point where the physician's decision might become nefarious. Another issue stems from the beauty of the beast itself: potentially it could learn and self-improve endlessly. Further data fed into the algorithm may improve the output again and again. Regulatory agencies will be put in the position of approving and possibly reimbursing tools that will continue to evolve after their marketing authorisation and actual clinical implementation. This is not a problem if we are sure that this evolution is for the better and if the self-improvement is sustained and not replaced by degradation in diagnostic performances. Continuous verification processes will need to be established, very much like quality controls performed on imaging and counting devices.

There will be a lengthy period of testing and validation, which will allow those of us who are willing to accept and promote those changes and learn the AI language and rules to play around with these new tools. This validation period and the early clinical implementation phase should be very exciting for everyone interested in the combination of new technologies and patient care, which includes most of us. Physicians must then enter the game, because progress in this field is currently advanced primarily by computer and other scientists, who are not trained in patient care.

The trick

Is everything really so neat and tidy? Most commentaries tend to embrace and integrate AI in the field of radiology and nuclear medicine and project an overall positive picture of AI entities deeply embedded in our daily practice, but under our control, much like a bright and reliable assistant [44–46]. It is true, that considering the enormous work represented by the validation process, which is mandatory for an AI algorithm to be integrated as a recognised clinical tool, this will take quite some time, at least if we do it the usual way: collecting data in single centres or in limited numbers of centres in the rare occurrences of multicentre studies, painstakingly tagging the imaging data with the appropriate clinical data in terms of endpoints, trying to also associate medical conditions that could influence the imaging results, and hopefully collecting other relevant data such as genomics, etc. All this should be performed with the utmost care and attention, because the quality of the input data (the ground truth) greatly impacts the reliability of the output. This is particularly true considering the current methodology: start with a narrow question, i.e., develop an algorithm to accomplish a narrow task with the purest cohort, validate it in a distinct but still very pure cohort, and finally test it in the real world. In order to achieve a comprehensive evaluation of the information contained in an imaging study, we would have to repeat the operation with multiple narrow questions and validate multiple narrow algorithms. This could give us the time and opportunity to learn how to “tame the beast” and maintain control of the process. We would use an increasing number of AI toolboxes the same way we use statistical parametric mapping in brain PET studies or comparison with normal databases in routine myocardial SPECT, only those toolboxes would be more potent and more reliable.

We may not be given this opportunity, however. AI as a whole is nothing new, but deep learning in medicine is less than 10 years old. We are at the very beginning of the process, and very few physicians, if any, have a clear view of all the capacities that deep learning will develop. As outlandish as it might seem, data mining of digitalised patients medical records could very well result in alleviating the need for the physician-driven testing and

validation process [47]. All agree, that patient care is based upon a relationship between the patient and the physician and characterised by transparency and informed decision making. The physician should be able to explain the pros and cons of options, possible outcomes, and eventually provide sound and understandable advice. This is the exact opposite of the black box effect, where we understand and choose what goes in (the input) and what comes out the AI model (the output), but cannot explain what goes on inside the box or exactly how the results are obtained. Common sense tells us that this black-box effect should be avoided at all cost, but this is not so simple. This notion is subject relative: most radiologists and nuclear medicine physicians are totally incapable of explaining how a single-layer neural network works, let alone deep-learning, and, consequently, they will describe it as a black box. Other scientists without medical knowledge may fully comprehend the model and could explain the intricacies of the algorithm, it is no black-box to them. If its validity is established, i.e., if the results lead to improved outcome in terms of quality of life or survival, who is to say that a patient will absolutely request a full explanation from their physician as to how the analysis was made? In any case, regarding diagnostic radiology and nuclear medicine procedures, this is not quite as occupationally relevant, because we often tend to interact more with clinicians than with the patient. The only issue left in this scheme is accountability, determining who is responsible for choosing and applying the medical management. Ultimately, the patient could very well complete their journey toward empowerment through a lengthy education process, guided by the barriers put in place by centralised regulators, and partly motivated by costs and other non-medical problems. Naturally, they would be accompanied by a smiling face in a white coat, who would lend the appropriate empathy.

As Niels Bohr is quoted to have said, “It’s difficult to make predictions, especially about the future”, so pick your own: the bright AI-augmented but physician-controlled future, or the dystopian nightmare. Facebook started with a noble idea, and has become a threat to democracy and individual freedom. One cannot discuss the current status of AI without mentioning the prospect, or spectre, of superintelligence, even though this remains a more philosophical than practical question.

Conclusion

AI is rapidly taking hold in all aspects of radiology and nuclear medicine. The potential benefits are obvious and enormous. Physicians still have the opportunity to accompany the changes, if not lead the process entirely. If fully designing and controlling the end-result seems an elusive goal, we can at the very least influence the way these changes are implemented.

Acknowledgements The author expresses his gratitude toward Dr. Nadia Withofs for fruitful discussion, and John Bean, for text editing.

Compliance with ethical standards

Conflict of interest The author has received a speaker honorarium from GE Healthcare, outside the scope of this manuscript.

There is no other conflict of interest.

References

1. Harvey HB, Liu C, Ai J, Jaworsky C, Guerrier CE, Flores E, et al. Predicting no-shows in radiology using regression modeling of data available in the electronic medical record. *J Am Coll Radiol*. 2017;14:1303–9.
2. Li X, Wang J, Fung RYK. Approximate dynamic programming approaches for appointment scheduling with patient preferences. *Artif Intell Med*. 2018;85:16–25.
3. Marella WM, Spamon E, Finley E. Screening electronic health record-related patient safety reports using machine learning. *J Patient Saf*. 2017;13:31–6.
4. Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *Neuroimage*. 2018;174:550–62.
5. Liu F, Jang H, Kijowski R, Zhao G, Bradshaw T, McMillan AB. A deep learning approach for (18)F-FDG PET attenuation correction. *EJNMMI Phys*. 2018;5:24.
6. Petersen H, Holdgaard PC, Madsen PH, Knudsen LM, Gad D, Gravergaard AE, et al. FDG PET/CT in cancer: comparison of actual use with literature-based recommendations. *Eur J Nucl Med Mol Imaging*. 2016;43:695–706.
7. Schmidt-Hansen M, Baldwin DR, Hasler E, Zamora J, Abairra V, Roque IFM. PET-CT for assessing mediastinal lymph node involvement in patients with suspected resectable non-small cell lung cancer. *Cochrane Database Syst Rev*. 2014;CD009519.
8. Helsen N, Van den Wyngaert T, Carp L, Stroobants S. FDG-PET/CT for treatment response assessment in head and neck squamous cell carcinoma: a systematic review and meta-analysis of diagnostic performance. *Eur J Nucl Med Mol Imaging*. 2018;45:1063–71.
9. Jaarsma C, Leiner T, Bekkers SC, Crijns HJ, Wildberger JE, Nagel E, et al. Diagnostic performance of noninvasive myocardial perfusion imaging using single-photon emission computed tomography, cardiac magnetic resonance, and positron emission tomography imaging for the detection of obstructive coronary artery disease: a meta-analysis. *J Am Coll Cardiol*. 2012;59:1719–28.
10. Huang JY, Huang CK, Yen RF, Wu HY, Tu YK, Cheng MF, et al. Diagnostic performance of attenuation-corrected myocardial perfusion imaging for coronary artery disease: a systematic review and meta-analysis. *J Nucl Med*. 2016;57:1893–8.
11. Nudi F, Iskandrian AE, Schillaci O, Peruzzi M, Frati G, Biondi-Zoccai G. Diagnostic accuracy of myocardial perfusion imaging with CZT technology: systemic review and meta-analysis of comparison with invasive coronary angiography. *JACC Cardiovasc Imaging*. 2017;10:787–94.
12. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast Cancer. *JAMA*. 2017;318:2199–210.
13. Nishio M, Sugiyama O, Yakami M, Ueno S, Kubo T, Kuroda T, et al. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. *PLoS One*. 2018;13:e0200721.

14. Artzi M, Bressler I, Ben Bashat D. Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. *J Magn Reson Imaging*. 2019. <https://doi.org/10.1002/jmri.26643>.
15. Gao X, Chu C, Li Y, Lu P, Wang W, Liu W, et al. The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from (18)F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. *Eur J Radiol*. 2015;84:312–7.
16. Blanc-Durand P, Van Der Gucht A, Schaefer N, Itti E, Prior JO. Automatic lesion detection and segmentation of 18F-FET PET in gliomas: a full 3D U-Net convolutional neural network study. *PLoS One*. 2018;13:e0195798.
17. Choi H, Jin KH. Alzheimer's disease neuroimaging I. predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behav Brain Res*. 2018;344:103–9.
18. Kim DH, Wit H, Thurston M. Artificial intelligence in the diagnosis of Parkinson's disease from ioflupane-123 single-photon emission computed tomography dopamine transporter scans using transfer learning. *Nucl Med Commun*. 2018;39:887–93.
19. Shibutani T, Nakajima K, Wakabayashi H, Mori H, Matsuo S, Yoneyama H, et al. Accuracy of an artificial neural network for detecting a regional abnormality in myocardial perfusion SPECT. *Ann Nucl Med*. 2019;33:86–92.
20. Cronin P, Dwamena BA, Kelly AM, Carlos RC. Solitary pulmonary nodules: meta-analytic comparison of cross-sectional imaging modalities for diagnosis of malignancy. *Radiology*. 2008;246:772–82.
21. Ruilong Z, Daohai X, Li G, Xiaohong W, Chunjie W, Lei T. Diagnostic value of 18F-FDG-PET/CT for the evaluation of solitary pulmonary nodules: a systematic review and meta-analysis. *Nucl Med Commun*. 2017;38:67–75.
22. 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.
23. Karantanis D, Kalkanis D, Czernin J, Hermann K, Pomykala KL, Bogsrud TV, et al. Perceived misinterpretation rates in oncologic 18F-FDG PET/CT studies: a survey of referring physicians. *J Nucl Med*. 2014;55:1925–9.
24. Wu AW, Cavanaugh TA, McPhee SJ, Lo B, Micco GP. To tell the truth: ethical and practical issues in disclosing medical mistakes to patients. *J Gen Intern Med*. 1997;12:770–5.
25. Pinto A, Brunese L, Pinto F, Reali R, Daniele S, Romano L. The concept of error and malpractice in radiology. *Semin Ultrasound CT MR*. 2012;33:275–9.
26. Degnan AJ, Ghobadi EH, Hardy P, Krupinski E, Scali EP, Stratchko L, et al. Perceptual and interpretive error in diagnostic radiology-causes and potential solutions. *Acad Radiol*. 2019;26(6):833–845. <https://doi.org/10.1016/j.acra.2018.11.006>.
27. Balint BJ, Steenburg SD, Lin H, Shen C, Steele JL, Gunderman RB. Do telephone call interruptions have an impact on radiology resident diagnostic accuracy? *Acad Radiol*. 2014;21:1623–8.
28. 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.
29. Iyer RS, Swanson JO, Otto RK, Weinberger E. Peer review comments augment diagnostic error characterization and departmental quality assurance: 1-year experience from a children's hospital. *AJR Am J Roentgenol*. 2013;200:132–7.
30. Wolf M, Krause J, Carney PA, Bogart A, Kurvers RH. Collective intelligence meets medical decision-making: the collective outperforms the best radiologist. *PLoS One*. 2015;10:e0134269.
31. Geijer H, Geijer M. Added value of double reading in diagnostic radiology, a systematic review. *Insights Imaging*. 2018;9:287–301.
32. Ulaner GA, Mannelli L, Dunphy M. Value of second-opinion review of outside institution PET-CT examinations. *Nucl Med Commun*. 2017;38:306–11.
33. Kuhl CK, Alparslan Y, Schmoee J, Sequeira B, Keulers A, Brummendorf TH, et al. Validity of RECIST version 1.1 for response assessment in metastatic cancer: a prospective, multireader study. *Radiology*. 2019;290:349–56.
34. Garcia EV, Klein JL, Moncayo V, Cooke CD, Del'Aune C, Folks R, et al. Diagnostic performance of an artificial intelligence-driven cardiac-structured reporting system for myocardial perfusion SPECT imaging. *J Nucl Cardiol*. 2018. <https://doi.org/10.1007/s12350-018-1432-3>.
35. Panayides AS, Pattichis M, Leandrou S, Pitris C, Constantinidou A, Pattichis CS. Radiogenomics for precision medicine with a big data analytics perspective. *IEEE J Biomed Health Inform*. 2018. <https://doi.org/10.1109/JBHI.2018.2879381>.
36. Pinker K, Chin J, Melsaether AN, Morris EA, Moy L. Precision medicine and radiogenomics in breast cancer: new approaches toward diagnosis and treatment. *Radiology*. 2018;287:732–47.
37. Jackson P, Hardcastle N, Dawe N, Kron T, Hofman MS, Hicks RJ. Deep learning renal segmentation for fully automated radiation dose estimation in unsealed source therapy. *Front Oncol*. 2018;8:215.
38. Kirienko M, Cozzi L, Rossi A, Voulaz E, Antunovic L, Fogliata A, et al. Ability of FDG PET and CT radiomics features to differentiate between primary and metastatic lung lesions. *Eur J Nucl Med Mol Imaging*. 2018;45:1649–60.
39. Hsu CY, Dubrovin M, Hua CH, Mohammed O, Shulkin BL, Kaste S, et al. Radiomics features differentiate between normal and tumoral high-Fdg uptake. *Sci Rep*. 2018;8:3913.
40. Deist TM, Dankers F, Valdes G, Wijnsman R, Hsu IC, Oberije C, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers. *Med Phys*. 2018;45:3449–59.
41. Callister ME, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax*. 2015;70(Suppl 2):ii1–ii54.
42. Herder GJ, van Tinteren H, Golding RP, Kostense PJ, Comans EF, Smit EF, et al. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. *Chest*. 2005;128:2490–6.
43. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med*. 2013;369:910–9.
44. Hainc N, Federau C, Stieltjes B, Blatow M, Bink A, Stippich C. The bright, artificial intelligence-augmented future of neuroimaging reading. *Front Neurol*. 2017;8:489.
45. Chan S, Siegel EL. Will machine learning end the viability of radiology as a thriving medical specialty? *Br J Radiol*. 2019;92:20180416.
46. Hall M. Artificial intelligence and nuclear medicine. *Nucl Med Commun*. 2019;40:1–2.
47. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep*. 2016;6:26094.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.