

Artificial Intelligence in Nuclear Medicine

Felix Nensa¹, Aydin Demircioglu¹, and Christoph Rischpler²

¹Department of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany; and ²Department of Nuclear Medicine, University Hospital Essen, University of Duisburg-Essen, Germany

Despite the great media attention for artificial intelligence (AI), for many health care professionals the term and the functioning of AI remain a “black box,” leading to exaggerated expectations on the one hand and unfounded fears on the other. In this review, we provide a conceptual classification and a brief summary of the technical fundamentals of AI. Possible applications are discussed on the basis of a typical work flow in medical imaging, grouped by planning, scanning, interpretation, and reporting. The main limitations of current AI techniques, such as issues with interpretability or the need for large amounts of annotated data, are briefly addressed. Finally, we highlight the possible impact of AI on the nuclear medicine profession, the associated challenges and, last but not least, the opportunities.

Key Words: artificial intelligence; machine learning; deep learning; nuclear medicine; medical imaging

J Nucl Med 2019; 60:29S–37S

DOI: 10.2967/jnumed.118.220590

In the field of medicine, in particular, medical imaging, the hype of recent years about artificial intelligence (AI) has had a significant impact. Although news in the daily press and medical publications about new capabilities and achievements of AI is almost overwhelming, for many interpreters the term and the functioning of AI remain a “black box,” leading to exaggerated expectations on the one hand and unfounded fears on the other. People already interact with AI in a variety of ways in everyday life—for example, on smartphones, in the car, or while surfing the internet—but often without actually realizing it. AI also has the potential to take on a variety of simple or repetitive tasks in the health care sector in the near future. However, AI certainly will not make radiologists or nuclear medicine specialists obsolete as medical experts in the foreseeable future. Rather than the disruption conjured up in some media, a steady transformation can be expected; this transformation most likely will begin or has begun in the diagnostic disciplines, in particular, medical imaging. From the perspective of the radiologist or nuclear medicine specialist, this development, instead of being perceived as a threat, can be seen as an opportunity to play a pioneering role within the health care sector and to actively shape this transformation process.

In this article, we attempt to provide a conceptual classification of AI, a brief summary of what we consider to be the most important technical fundamentals, a discussion of possible applications in nuclear medicine and, finally, a brief consideration of the possible impact of these technologies on the profession of the physician.

HOW TO DEFINE AI

The term *artificial intelligence* first appeared in an application for a 6-wk workshop entitled Dartmouth Summer Research Project on Artificial Intelligence at Dartmouth College in Hanover, New Hampshire (1), and is often defined as “intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and animals” (Wikipedia; https://en.wikipedia.org/wiki/Artificial_intelligence). However, since its first appearance, the term has undergone a constant redefinition against the background of what is technically feasible. On the one hand, the definition per se is already vague, because the partial term *intelligence* is not itself well defined. On the other hand, it depends directly on human perception and evaluation, which change constantly. Only a few decades ago, chess computers were regarded as a classic example of AI, because a kind of “intelligence” was considered a prerequisite for the ability to master this game. With the exponential growth of performance in computer hardware, however, it was soon possible to program chess computers that played masterfully without developing an understanding of the game as human players do. In simple terms, a computer’s memory had stored such a large number of moves from archived chess games between professional human players that the computer could look up an equivalent in a historical game for almost every imaginable game situation and derive the next move from it. This procedure, although simplified here, did produce extremely successful chess computers, but their behavior was predictable in principle and lacked typical human qualities, such as strategy and creativity. This “explainability,” together with a certain wear and tear of the “wow effect,” finally led to the fact that chess computers are no longer regarded as examples of AI by most people today.

An attempt to systematize the area of AI leads to a multitude of different procedures which, only in their entirety, define the field of AI (Figs. 1 and 2). From the 1950s to the 1980s, AI was strongly dominated by so-called symbolic reasoning, through which AI is implemented by rules engines, expert systems, or so-called knowledge graphs. What these methods have in common is that they model entities of the real world and their logical relationships in the form of symbols with which arithmetic operations can then be performed. The main advantages of these systems are, on the one hand, their often comparatively low demand on the computing capacity of a computer system and, on the other

Received Apr. 1, 2019; revision accepted May 16, 2019.

For correspondence or reprints contact: Felix Nensa, Department of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Essen, University of Duisburg-Essen, Hufelandstrasse 55, 45147 Essen, Germany.

E-mail: felix.nensa@uk-essen.de

COPYRIGHT © 2019 by the Society of Nuclear Medicine and Molecular Imaging.

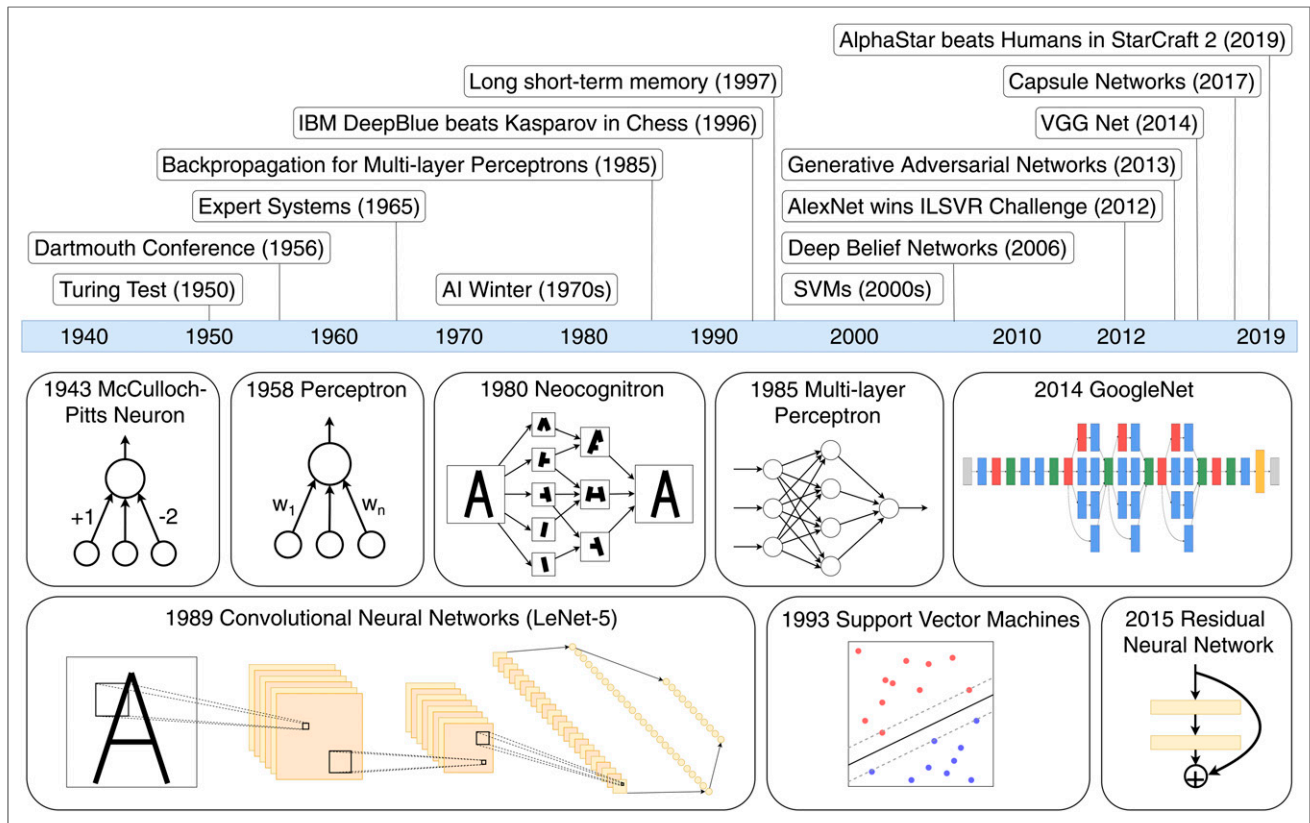


FIGURE 1. Brief time line of major developments in AI and machine learning. Some methods are also depicted symbolically. ILSVR = ImageNet Large Scale Visual Recognition; SVMs = support vector machines; VGG = Visual Geometry Group.

hand, their comprehensible behavior, with which every step of the system (data input, processing, and data output) can be reproduced and understood. The main disadvantage, however, is the necessary step of modeling, in which the part of the real world required for the concrete application domain has to be converted into symbols. This extremely labor-intensive task often has to be performed by people, so that the creation of such systems is mostly reserved for corporations (e.g., Google; <https://www.google.com/bn/search/about/>) or, recently, well-organized crowdsourcing movements (e.g., Wikidata; <https://www.wikidata.org>).

The larger problem in modeling, however, is that the performance and accuracy of such systems are bound a priori to the human

understanding of the real world. Although this situation seems unproblematic in a game such as chess, with a manageable number of game pieces and their well-defined relationships to each other, for other applications (such as medicine), this situation results in considerable difficulties. Thus, many physicians are probably aware that even the most complex medical ontologies and classifications ultimately represent crude simplifications of the underlying biologic systems and do not fully describe the variability of diseases or their dependencies. Moreover, such classification systems can hardly keep pace with the medical knowledge gained in the digital age, a fact that inevitably limits symbolic AI systems based on such models.

However, with the strongly increasing performance of computer hardware, nonsymbolic AI systems increasingly came to the fore from the mid-1980s onward. What these systems have in common is that they are data driven and work statistically. These procedures are often summarized under the term *machine learning*, in which computer systems learn to accomplish a task independently—that is, without explicit instructions—and thus perform observational learning from large amounts of data. The obvious advantage of these systems is that the time-consuming and limiting modeling phase is omitted, because the machine largely independently appropriates the internal abstraction of the respective problem and, assuming a sufficient and representative amount of example data, can also record and map its variability. In addition to the high demand for computing capacity during the training phase, these methods primarily have 2 disadvantages. On the one hand, there is a large to very large demand for example datasets during the training

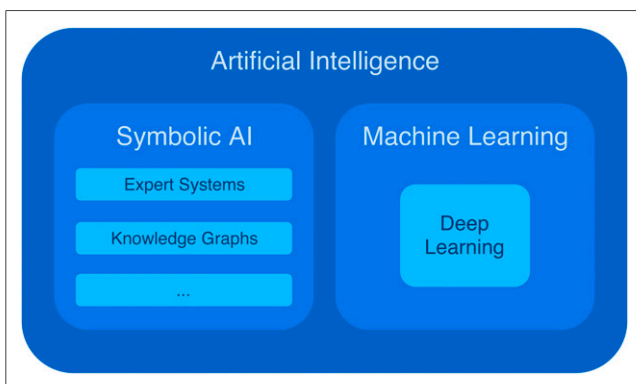


FIGURE 2. Division of field of AI into symbolic AI and machine learning, of which deep learning is a branch.

phase for almost all methods because, despite all technical advances, the abstraction of a problem is far less efficient than in the human brain. On the other hand, the internal representation of this abstraction in most of these systems is so complex that it can no longer be comprehended and understood by people, so that such systems are often referred to as “black boxes,” and the corresponding output of such systems can no longer be reliably predicted outside the set of tested input parameters. For complex and highly variable input parameters, such as medical image data, these systems thus can produce unexpected results and show a quasi-nondeterministic behavior; for example, an image of an elephant can be placed clearly visible into an image, and a state-of-the-art trained neural network either will most often not see it at all or will mistake it as other objects, such as a chair (2).

In principle, machine learning procedures can be divided into supervised and unsupervised learning. In supervised learning, not only the input data but also the desired output data are given during the training phase, and the model learns to generate those outputs from the given inputs. To prevent the model from learning only the example data by memorization (also referred to as overfitting), various techniques are used; the central element is that only part of the data is presented to the model during training, and the performance of the model (i.e., the control of learning success) is measured against the other part of the data. In contrast, in unsupervised learning, the input data are given without any labels. The goal is then to understand the inherent structure in the data. Using clustering methods, for example, the observations to be analyzed are divided into subgroups according to certain features or feature combinations. Generative methods derive a probability distribution from sampled observations that can be used to generate synthetic observations. In the medical domain, in which the cost of labeling the data is high, semisupervised learning could be more useful. Here, only part of the data is labeled, and although the task is similar to supervised learning, the advantage is that the structure of the unlabeled data—which are often more abundant—can be exploited.

Another form of classification is the division of the area of machine learning into conventional machine learning and deep learning. Conventional machine learning includes a large number of established methods, such as naive Bayes classifiers, support vector machines, random forests, or even hidden Markov models, and has been used for years and decades in a wide variety of application areas, such as time series predictions, recommendation engines in e-commerce, spam filters, text translation, and many more. In recent years, however, the field of machine learning has been strongly influenced by deep learning, which is based on artificial neural networks (ANNs). Because of a multitude of layers (so-called hidden layers) between the input and output layers, these neural networks have a much larger space for free parameters and thus allow much more complex abstractions than conventional machine learning methods.

An area of medical imaging currently receiving much attention, so-called radiomics, can be reduced to a 2-step process. In the first step, image data are converted by image processing methods into high-dimensional vectors (so-called feature vectors); from these vectors, predictive models—usually a classifier or a regressor—for deriving certain information from the same image data are then generated in the second step using conventional machine learning. Radiomics is currently being evaluated in a multitude of small, often retrospective studies, which often try to predict information such as histologic subtype, mutational status, or a response to a

certain therapy from medical images of tumors. Because the first step requires careful feature engineering and strong domain expertise, there are already some attempts to replace the 2-step process in radiomics with deep learning by placing the image data directly into the input layer of an ANN without prior feature extraction. Because an article dedicated to radiomics also appears in this supplement to *The Journal of Nuclear Medicine*, we will not discuss radiomics further and will focus in particular on other applications of machine learning and deep learning (Supplemental Appendix 1) (supplemental materials are available at <http://jnm.snmjournals.org>) (3–7).

APPLICATIONS IN NUCLEAR MEDICINE

The rise of AI in medicine is often associated with “superhuman” abilities and precision medicine. At the same time, often overlooked are the facts that large parts of physicians’ everyday work consist of routine tasks and that the delegation of those tasks to AI would give the human workforce more time for higher-value activities (8) that typically require human attributes such as creativity, cognitive insight, meaning, or empathy. The day-to-day work of medical imaging involves a multitude of activities, including the planning of examinations, the detection of pathologies and their quantification, and manual research for additional information in medical records and textbooks—which often tend to bore and demand too little intellectually from the experienced physician but, with continuously rising workloads, tend to overwhelm the beginner. Without diminishing the prospects of “superdiagnostics” and precision medicine, seemingly more easily achievable goals of AI in medicine should not be forgotten because they might relieve people who are highly educated and have specialized skills of repetitive routine tasks.

A typical medical imaging work flow can be divided into 4 steps: planning, image acquisition, interpretation, and reporting (Fig. 3). Steps such as admission and payment could be included as well. We have deliberately focused on the parts of the work flow in which the physician is directly and primarily involved. In Figure 3, each step is assigned a list with examples of typical tasks that could be performed in that step and that could be improved, accelerated, or completely automated with the help of AI. Next, we discuss existing or potential AI-based solutions clustered by that structure.

Planning

Before an examination is performed on a patient at all, whether a planned procedure is medically indicated should be determined. The more unpleasant, risky, or expensive the respective examination is, the more this guideline applies. For example, in the recruitment of amyloid-positive individuals for clinical trials, Ansart et al. showed that screening based on conventional machine learning with random forests and cognitive, genetic, and sociodemographic features led to an increased number of recruitments and to a reduced number of costly (~€1,000 in Europe and \$5000 in the United States) PET scans (9).

One of the greatest challenges in the scheduling of medical examinations is “no-shows”; this challenge is particularly problematic in the case of nuclear medicine because of tracer availability, decay, and cost. A highly relevant study from Massachusetts General Hospital demonstrated the feasibility of predicting no-shows in the medical imaging department using relatively simple machine learning algorithms and logistic regression (10). The authors included 54,652 patient appointments with scheduled radiology examinations in their

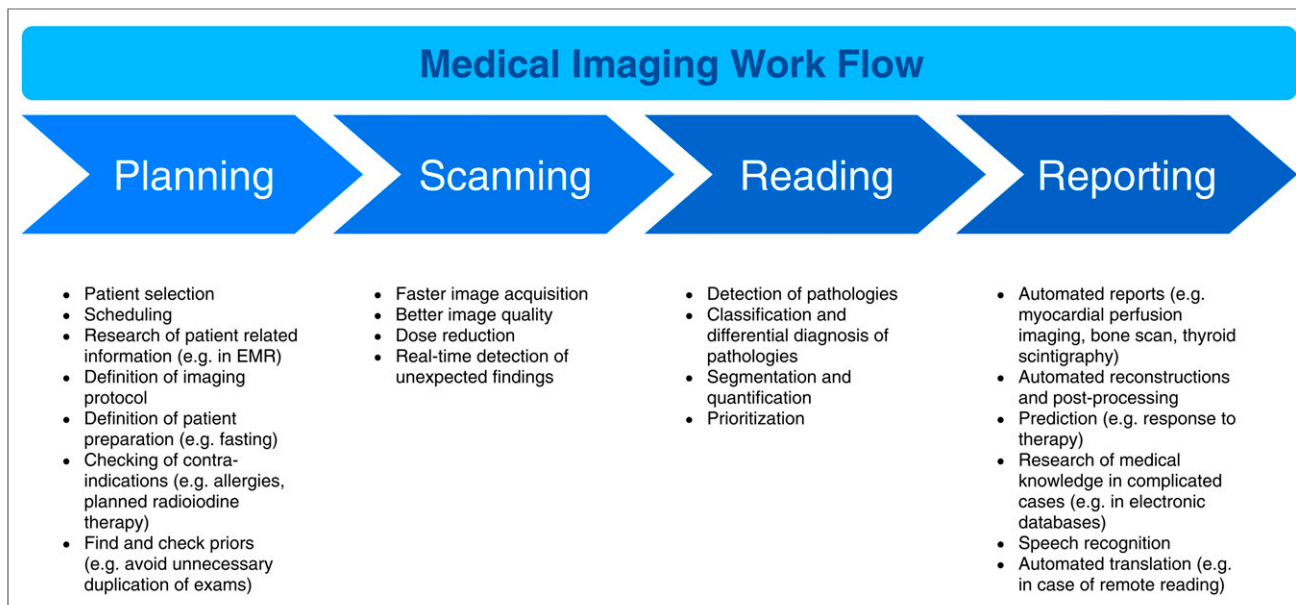


FIGURE 3. Division of typical medical imaging Work Flow into 4 steps: planning, image acquisition, interpretation (reading), and reporting. Each step is assigned a list with examples of typical tasks that could be performed in that step and could be improved, accelerated, or completely automated with the help of AI. EMR = electronic medical record.

study. Considering 16 data elements from the electronic medical record grouped by no-show history, appointment-specific factors, and sociodemographic factors, their model had a significant power to predict failure to attend a scheduled radiology examination (area under the curve [AUC], 0.75) (10). Given the recent technical improvements in deep learning, the relatively small number of included predictors in that study, and the recent availability of methods such as continuous (or incremental) learning, it is not far-fetched to hypothesize that the prediction of no-shows at a much higher accuracy could be available soon.

Often patient-related information given at the time of referral is sparse, and extensive manual searching through large numbers of unstructured text documents by the physician is necessary to gather all of the information that is needed for optimal preparation and planning of the examination. Although the analysis of text documents may seem easy (compared with, e.g., image analysis) and recent advances in natural language processing and natural language understanding became very visible in gadgets such as Alexa (<https://alexa.amazon.com>), Google Assistant (<https://assistant.google.com>), or Siri (<https://www.apple.com/siri/>), such analysis in fact remains a particularly delicate task for machine learning. Still, the research community is making steady progress (11), structured reporting that allows straightforward algorithmic information extraction is gaining popularity (12), and data interoperability standards such as Fast Healthcare Interoperability Resources (FHIR) (<https://www.hl7.org/fhir/>) will gradually become available in clinical systems. Therefore, it can be assumed that, in the future, the time-consuming manual research of patient information will be performed by intelligent artificial assistants and presented to the physician in the form of concise case-specific dashboards. Such dashboards not only will aggregate relevant patient information but also likely will enrich this information by putting it into context. For example, a relatively simple rule-based symbolic AI could automatically check for certain contraindications, such as allergies, or reduce unnecessary duplication of examinations by analyzing prior examinations.

Scanning

Modern scanner technology already makes increasing use of machine learning, and recent advancements in research suggest considerable technical improvements in the near future (13). In nuclear medicine, attenuation maps and scatter correction remain hot topics for PET and SPECT imaging, so it is not surprising that these are the subjects of intensive research by various AI groups. Hwang et al. used a modified U-Net, which is a specialized convolutional network architecture for biomedical image segmentation (14), to generate the attenuation maps for whole-body PET/MRI (15). They used activity and attenuation maps estimated from the maximum-likelihood reconstruction of activity and attenuation algorithm as inputs to create a CT-derived attenuation map and compared this method with the Dixon-based 4-segment method. Compared with the CT-derived attenuation map, the U-Net-based approach achieved significantly higher agreement (Dice coefficient, 0.77 vs. 0.36). Instead of an analytic approach based on image segmentation, it is also possible to use generative adversarial networks (GANs) to directly translate 1 imaging modality into another. The feasibility of direct MR-to-CT image translation using context-aware GANs was demonstrated by Nie et al. in a small study involving 15 brain and 22 pelvic examinations (16).

Another topic of research is the improvement of image quality. Hong et al. used a deep residual convolutional neural network (CNN) to enhance the image resolution and noise property of PET scanners with large pixelated crystals (17). Kim et al. showed that iterative PET reconstruction using a denoising CNN with local linear fitting improved image quality and was robust against noise-level disparities (18). Improvements in reconstructed image quality could also be translated to dose savings, as shown by multiple groups that estimated full-dose PET images from low-dose scans (i.e., reduction in applied radioactivity) using CNNs (19,20) or GANs (21) with favorable results. Obviously, this approach could also be translated to shorter acquisition times and result in higher patient throughput. In addition, improved image

quality could also be translated to higher temporal resolution, as shown by Cui et al., who used stacked sparse autoencoders (unsupervised ANNs that learn a representation by training the network to ignore noise) to improve the quality of dynamic PET images (22). Berg and Cherry used CNNs to estimate time-of-flight directly from the pair of digitized detector waveforms for a coincident event; this method improved timing resolution by 20% compared with leading-edge discrimination and 23% compared with constant fraction discrimination (23). An interesting approach was pursued in the study of Choi et al. (24). There, virtual MR images generated from florbetapir PET images using GANs were then used for quantification of the cortical amyloid load (mean \pm SD absolute error of the SUV ratio of cortical composite regions, 0.04 ± 0.03); in principle, this method could make the additional MRI scan obsolete.

In nuclear medicine, scanning depends directly on the application of radiotracers, the development of which is a time-consuming and costly process. As in the pharmaceutical industry, the prediction of drug–target interactions (DTI) is an important part of this process in the radiopharmaceutical industry and has been performed with computer assistance for quite some time; AI-based methods are increasingly being used (25,26). For example, Wen et al. were able to predict the interactions between ziprasidone or clozapine and the 5-hydroxytryptamine receptor 1C (or 2C) or alprazolam and γ -aminobutyric acid receptor subunit ρ -2 with a deep-belief network (25).

Interpretation

Many interpreters maintain a list of examinations that they have to interpret and that they process chronologically in a first-in, first-out order. In reality, however, some studies have findings that require prompt action and therefore should be prioritized. Recently, a deep learning–based triage system that detects free gas, free fluid, or fat stranding in abdominal CTs was published (27), and multiple studies have already demonstrated the feasibility of detecting critical findings in head CT scans (28–31). In the future, such systems could work directly on raw data, such as sinograms, and raise alerts during the scan time, even before reconstruction. In such a scenario, the technician could modify or extend the planned scan protocol to accommodate the unexpected finding; for example, an intracranial hemorrhage detected during a low-dose PET/CT scan could trigger an immediate full-dose CT scan of the head. However, the automatic detection of pathologies also offers other interesting possibilities beyond the prioritization of studies. For example, the processing of certain examinations, such as bone or thyroid scans, could be automated or at least accelerated with preliminary assessments, or an AI assistant working in the background could alert the human interpreter to possibly overlooked findings. Another, often disregarded possibility is that recurring secondary findings could be automatically detected and included in the report, freeing the human interpreter from an often annoying task.

Many studies have already addressed the early detection of Alzheimer disease and mild cognitive impairment using deep learning (32–37). Ding et al. were able to show that a CNN with InceptionV3 architecture (38) could make an Alzheimer disease diagnosis with 82% specificity at 100% sensitivity (AUC, 0.98) on average 75.8 mo before the final diagnosis based on ^{18}F -FDG PET/CT scans and outperformed human interpreters (majority diagnosis of 5 interpreters) (39). A similar network architecture was used by Kim et al. in the diagnosis of Parkinson disease from

^{123}I -ioflupane SPECT scans; the test sensitivity was 96.3% at 66.7% specificity (AUC, 0.87) (40). Li et al. used a 3-step process of automatic segmentation, feature extraction, and classification using support vector machines and random forests to automatically detect pancreas carcinomas on ^{18}F -FDG PET/CT scans (41). On their test dataset of 80 scans, they found a sensitivity of 95.23% at a specificity of 97.51% (41). Perk et al. combined threshold-based detection with machine learning–based classification to automatically evaluate ^{18}F -NaF PET/CT scans for bone metastases in patients with prostate cancer (42). A combination of statistically optimized regional thresholding and random forests resulted in a sensitivity of 88% at a specificity of 89% (AUC, 0.95) (42). However, the ground truth in learning data originated from only 1 human interpreter, so that the performance of the machine learning approach must be evaluated with care. Interestingly, in a subset of patients who were evaluated by 3 additional nuclear medicine specialists, the machine learning classification performance was high when the ground truth originated from any of the 4 physicians (AUC range, 0.91–0.93), whereas the agreement between the physicians was only moderate (κ , 0.53). That study (42) underlined the importance of reliable ground truth not only during validation but also during training when supervised learning is used. Nevertheless, it should not be forgotten that although existing systems sometimes provide excellent results with regard to the detection of 1 or more classes of pathologies, they still cannot generalize results as well as a human diagnostician. For this reason, human supervision remains absolutely mandatory in most scenarios.

Overall, however, the detection of pathologies during interpretation often accounts for only a small part of the total effort for the experienced interpreter. The increasing demand for quantification and segmentation usually involves much more effort, although these tasks are intellectually not very challenging and often are rather tiring. Therefore, the reasons for the wish to delegate these tasks to intelligent systems seem obvious. Rocca et al. used machine learning to estimate the arterial input function for the non-invasive full quantification of the regional cerebral metabolic rate for glucose in ^{18}F -FDG PET (43). Instead of measuring the arterial input function during the scan with an invasive arterial blood sampling procedure, it was predicted with data from medical health records and dynamic PET imaging data. Before planned radiotherapy, it is necessary to precisely quantify the target structures by segmentation which, in the case of nasopharyngeal carcinomas, is often a particularly difficult and time-consuming activity because of the anatomic location. Zhao et al. showed, for a small group of 30 patients, that the automatic segmentation of such tumors on ^{18}F -FDG PET/CT data was, in principle, possible using the U-Net architecture (mean Dice score of 87.47%) (44). Other groups applied similar approaches to head and neck cancer (45) and lung cancer (46,47). Still, fully automated tumor segmentation remains a challenge, probably because of the extremely diverse appearance of these diseases. Such an approach requires correspondingly large amounts of training data, for which the necessary ground truth in the form of segmentation masks usually has to be generated in a labor-intensive manual or semi-automatic task.

Intelligent systems can also support the interpreter with classification and differential diagnosis. Many studies have shown possible applications for radiology, such as the differentiation of liver masses in MRI (48), bone tumor diagnosis in radiography (49), classification of interstitial lung diseases in CT (50), or diagnosis of acute infarctlike myocarditis in MRI (51).

Togo et al. showed that the evaluation of polar maps from ^{18}F -FDG PET scans using deep learning for the presence of cardiac sarcoidosis yielded significantly better results (83.9% sensitivity at 87% specificity) than methods based on SUV_{max} (46.8% sensitivity at 71% specificity) or variance (65.5% sensitivity at 75% specificity) (52). Ma et al. used a modified DenseNet architecture pretrained by ImageNet to diagnose Graves disease, Hashimoto disease, and subacute thyroiditis on thyroid SPECT scans (53). The training dataset was considerably large, including 780 samples of Graves disease, 438 samples of Hashimoto disease, 810 samples of subacute thyroiditis, and 860 samples of normal cases. However, their validation strategy remains unclear, so the reported numbers must be evaluated with care (53).

Reporting

Medical imaging professionals are often confronted with referential questions that, according to current knowledge and the state of the art, cannot be answered reliably or at all with the possibilities of imaging. In health care, AI is often intuitively associated with superhuman performance, so it is not surprising that there is such a high level of research activity in the area of prediction of unknown outcomes.

Despite the high sensitivity and specificity of procedures such as PET/CT in tumor detection, it is still not possible to detect so-called micrometastases or early metastatic disease, although the detection of tumor spread has significant effects on the treatment concept. In an animal study of 28 rats injected with breast cancer cells, Ellmann et al. were able to predict later skeletal metastasis with an ANN based on ^{18}F -FDG PET/CT and dynamic contrast-enhanced MRI data on day 10 after injection with an accuracy of 85.7% (AUC, 0.90) (54). Future prospective studies will show whether these results can also be achieved in people, but the approach seems promising. Another group achieved promising results in the detection of micrometastases in lymph nodes in head and neck cancers by combining radiomics analysis of CT data and 3-dimensional CNN analysis of ^{18}F -FDG PET data through evidential reasoning (55).

Another important question in oncology—one that often cannot be answered with imaging—is the prediction of the response to therapy and overall survival. A small study by Xiong et al. of 30 patients with esophageal cancer demonstrated the feasibility of predicting local disease control with chemoradiotherapy using radiomics features from ^{18}F -FDG PET/CT and machine learning models (56). Milgrom et al. analyzed ^{18}F -FDG PET scans of 251 patients with stage I or II Hodgkin lymphoma (57). They found that 5 features extracted from mediastinal sites were highly predictive of primary refractory disease when incorporated into a machine learning model (57). In a study conducted to predict overall survival in glioblastoma multiforme by integrating clinical, pathologic, semantic MRI-based, and O -(2- ^{18}F -fluoroethyl)-L-tyrosine PET/CT-derived information as well as treatment features into a machine learning model, PET/CT was not found to provide additional predictive power; however, the fraction of patients with available PET data was relatively low (68/189), and 2 different PET reconstruction methods were used (58). A study by Papp et al. included L-S-methyl- ^{11}C -methionine PET features, histopathologic features, and patient characteristics in a machine learning model to predict 36-mo survival in 70 patients with treatment-naïve gliomas; an AUC of up to 0.9 was achieved (59). Ingrisch et al. tried to predict the outcome of ^{90}Y radioembolization in patients with intrahepatic tumors from pretherapeutic baseline parameters (60). They trained a random

survival forest with baseline levels of cholinesterase and bilirubin, type of primary tumor, age at radioembolization, hepatic tumor burden, presence of extrahepatic disease, and sex. Their model achieved a moderate predictive power, with a concordance index of 0.657, and identified baseline cholinesterase and bilirubin as the most important variables (60).

Reporting in nuclear cardiology often involves the prediction of coronary artery disease and the associated risk of major adverse cardiac events. A multicenter study of 1,160 patients without known coronary artery disease was conducted to evaluate the prediction of obstructive coronary disease from a combined analysis of semiupright and supine stress $^{99\text{m}}\text{Tc}$ -sestamibi myocardial perfusion imaging by a CNN versus a standard combined total perfusion deficit (61). To approximate external validation, the authors performed training using a leave-1-center-out cross-validation procedure. The AUC for the prediction of disease on a per-patient basis and a per-vessel basis was higher for the CNN than for the combined total perfusion deficit (per-patient AUC, 0.81 vs 0.78; per-vessel AUC, 0.77 vs 0.73) (61). The same group also evaluated the added predictive value of combining clinical information and myocardial perfusion imaging using the LogitBoost algorithm to predict major adverse cardiac events. They included a total of 2,619 consecutive patients and found that their model predicted a 3-y risk of major adverse cardiac events with an AUC of 0.81 (62).

Finally, when complex cases or rare diseases are being reported, it is often helpful to compare them with similar cases from databases and case collections. Although a textual search—for example, in archived reports—is uncomplicated, an image-based search is often not possible. Through AI-based automatic image annotations (63) and content-based image retrieval (64), conducting large, direct image-based and ad hoc database searches and thereby finding potentially similar cases that might be helpful in a real diagnostic situation are increasingly possible.

LIMITATIONS OF AI

Although the use of AI in health care certainly holds great potential, its limitations also need to be acknowledged. A well-known problem is the interpretability of the models. Although symbolic AI or simple machine learning models, such as decision trees or linear regression, are still fully understood by people, understanding becomes increasingly difficult with more advanced techniques and is now impossible with many deep learning models; this situation can lead to unexpected results and non-deterministic behavior (2). Although this issue also applies to other procedures in medicine in which the exact mechanisms of action are often poorly understood (e.g., pharmacotherapy), whether predictive AI can and may be used for far-reaching decisions if the exact mode of action is unclear remains unresolved. However, in cases in which AI acts as an assistant that provides hints or produces results that can be replicated by people or visually verified (e.g., by volumetry), the lack of interpretability of the underlying models may not be an obstacle to clinical application. For other cases, especially in image recognition and interpretation, certain techniques (such as activation maps) can provide high-level visual insights into the inner workings of ANNs (Fig. 4). The problem of interpretability is the subject of intensive research and various initiatives (65,66), although whether these will be able to keep pace with the rapid progress in the development of increasingly complex ANN architectures is unclear.

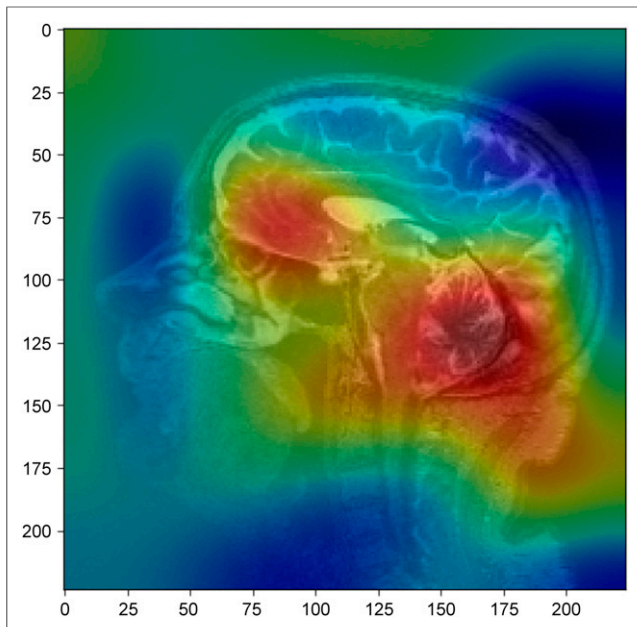


FIGURE 4. Sagittal T2-weighted reconstruction of brain MRI scan overlaid with activation map. This example is taken from training dataset for automatic detection of dementia using MRI scans with CNN. Activations show that CNN focuses strongly on frontobasal brain region and cerebellum for prediction. (Image courtesy of Obioma Pelka, Essen, Germany.)

Another problem is that many machine learning applications will always deliver a result on an input but cannot provide a measure of the certainty of their prediction (67). Thus, a human operator often cannot decide whether to trust the result of AI-based software or not. Possible solutions for this problem are the integration of probabilistic reasoning and statistical analysis in machine learning (68) as well as quality control (69). Bias and prejudice are well-known problems in medicine (70). However, training AI systems with biased data will make the resulting models generate biased predictions as well (71,72); this issue is especially problematic because many users perceive such systems as analytically correct and unprejudiced and therefore tend not to question their predictions in terms of bias. One of the largest hurdles for AI in health care is the need for large amounts of structured and annotated data for supervised learning. Many studies therefore work with small datasets, which are accompanied by overfitting and poor generalizability and reproducibility. Therefore, increased collaboration and standardization are needed to generate large machine-readable datasets that reflect variability in real populations and that have as little bias as possible.

OUTLOOK AND FUTURE PERSPECTIVE

Many publications on the topic of AI in medicine deal with some degree of automation. Whether it is the measurement (quantification and segmentation) of pathologies, the detection of pathologies, or even automated diagnosis, AI does not necessarily have to be superhuman to have a benefit for medicine. However, it is obvious that AI is already better than people in some areas, and this development is a consequence of technologic progress. Therefore, many physicians are concerned that they will be replaced by AI in the future—a concern that is partly exacerbated by insufficient knowledge of how AI works. On the other hand, Geoffrey Hinton,

undoubtedly one of the most renowned AI experts, made the statement, “People should stop training radiologists now!” at a conference in 2016 (73). This statement triggered a lot of contradiction (74–76) and is perhaps best explained by a lack of understanding of medicine in general and medical imaging in particular on his part.

Although most experts and surveys reject the fear of AI replacing physicians (77–79), this fact does not mean that AI will have no impact on the medical profession. In fact, it is highly likely that AI will transform the medical profession and medical imaging in particular. In the near future, the automation of labor-intensive but cognitively undemanding tasks, such as image segmentation or finding prior examinations across different PACS repositories, will be available for clinical application. This change should be perceived not as a threat but as an opportunity to relieve oneself of this work and as a stimulus for the further development of the profession. In fact, it is imperative for the profession to grow into the role it will be given in the future by AI. The increasing use of large amounts of digital data in medicine will create the need for new skills, such as clinical data science, computer science, and machine learning, especially in diagnostic disciplines. It can even be assumed that the boundaries between the diagnostic disciplines will become blurred, as the focus will increasingly be less on the detection and classification of individual findings and more on the comprehensive analysis and interpretation of all available data on a patient (80). Although prospective physicians can be confident that medical imaging offers them a bright future, it is important for them to understand that this future is open only to those who are willing to acquire competencies like those mentioned earlier. Without the training of and necessary expertise among physicians, precision health care, personalized medicine, and superdiagnostics are unlikely to become clinical realities. As Chan and Siegel (77) and others have stated, physicians will not be replaced by AI, but physicians who opt out from AI will be replaced by others who embrace it.

DISCLOSURE

Felix Nensa is an academic collaborator with Siemens Healthineers and GE Healthcare. No other potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENT

Most of this work was written by the authors in German and translated into English using a deep learning–based web tool (<https://www.deepl.com/translator>).

REFERENCES

1. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Mag.* 2006;27:12–14.
2. Rosenfeld A, Zemel R, Tsotsos JK. The elephant in the room. arXiv.org website. <https://arxiv.org/abs/1808.03305>. Accessed June 20, 2019.
3. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. 1943. *Bull Math Biol.* 1990;52:99–115.
4. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev.* 1958;65:386–408.
5. Minsky M, Papert S. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press; 1969.
6. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: *Proceedings of the 25th International Conference*

- on Neural Information Processing Systems. Vol 1. Red Hook, NY: Curran Associates Inc.; 2012:1097–1105.
7. Korkinof D, Rijken T, O'Neill M, Yearsley J, Harvey H, Glocker B. High-resolution mammogram synthesis using progressive generative adversarial networks. arXiv.org website. <https://arxiv.org/abs/1807.03401>. Accessed June 20, 2019.
 8. 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.
 9. Ansart M, Epelbaum S, Gagliardi G, et al. Reduction of recruitment costs in preclinical AD trials: validation of automatic pre-screening algorithm for brain amyloidosis. *Stat Methods Med Res*. January 30, 2019 [Epub ahead of print].
 10. Harvey HB, Liu C, Ai J, et al. Predicting no-shows in radiology using regression modeling of data available in the electronic medical record. *J Am Coll Radiol*. 2017;14:1303–1309.
 11. Pons E, Braun LMM, Hunink MGM, Kors JA. Natural language processing in radiology: a systematic review. *Radiology*. 2016;279:329–343.
 12. Pinto Dos Santos D, Baeßler B. Big data, artificial intelligence, and structured reporting. *Eur Radiol Exp*. 2018;2:42.
 13. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature*. 2018;555:487–492.
 14. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Cham, Switzerland: Springer International Publishing; 2015:234–241.
 15. Hwang D, Kang SK, Kim KY, et al. Generation of PET attenuation map for whole-body time-of-flight ¹⁸F-FDG PET/MRI using a deep neural network trained with simultaneously reconstructed activity and attenuation maps. *J Nucl Med*. January 25, 2019 [Epub ahead of print].
 16. Nie D, Trullo R, Lian J, et al. Medical image synthesis with context-aware generative adversarial networks. *Med Image Comput Comput Assist Interv*. 2017;10435: 417–425.
 17. Hong X, Zan Y, Weng F, Tao W, Peng Q, Huang Q. Enhancing the image quality via transferred deep residual learning of coarse PET sinograms. *IEEE Trans Med Imaging*. 2018;37:2322–2332.
 18. Kim K, Wu D, Gong K, et al. Penalized PET Reconstruction using deep learning prior and local linear fitting. *IEEE Trans Med Imaging*. 2018;37:1478–1487.
 19. Xiang L, Qiao Y, Nie D, An L, Wang Q, Shen D. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. *Neurocomputing*. 2017;267:406–416.
 20. Kaplan S, Zhu YM. Full-dose PET image estimation from low-dose PET image using deep learning: a pilot study. *J Digit Imaging*. November 6, 2018 [Epub ahead of print].
 21. Wang Y, Yu B, Wang L, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *Neuroimage*. 2018;174:550–562.
 22. Cui J, Liu X, Wang Y, Liu H. Deep reconstruction model for dynamic PET images. *PLoS One*. 2017;12:e0184667.
 23. Berg E, Cherry SR. Using convolutional neural networks to estimate time-of-flight from PET detector waveforms. *Phys Med Biol*. 2018;63:02LT01.
 24. Choi H, Lee DS; Alzheimer's Disease Neuroimaging Initiative. Generation of structural MR images from amyloid PET: application to MR-less quantification. *J Nucl Med*. 2018;59:1111–1117.
 25. Wen M, Zhang Z, Niu S, et al. Deep-learning-based drug-target interaction prediction. *J Proteome Res*. 2017;16:1401–1409.
 26. Chen R, Liu X, Jin S, Lin J, Liu J. Machine learning for drug-target interaction prediction. *Molecules*. 2018;23:2208.
 27. Winkel DJ, Heye T, Weikert TJ, Boll DT, Stieltjes B. Evaluation of an AI-based detection software for acute findings in abdominal computed tomography scans: toward an automated work list prioritization of routine CT examinations. *Invest Radiol*. 2019;54:55–59.
 28. Prevedello LM, Erdal BS, Ryu JL, et al. Automated critical test findings identification and online notification system using artificial intelligence in imaging. *Radiology*. 2017;285:923–931.
 29. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet*. 2018; 392:2388–2396.
 30. Majumdar A, Brattain L, Telfer B, Farris C, Scalera J. Detecting intracranial hemorrhage with deep learning. *Conf Proc IEEE Eng Med Biol Soc*. 2018;2018: 583–587.
 31. Cho J, Park KS, Karki M, et al. Improving sensitivity on identification and delineation of intracranial hemorrhage lesion using cascaded deep learning models. *J Digit Imaging*. 2019;32:450–461.
 32. Yamashita AY, Falcão AX, Leite NJ; Alzheimer's Disease Neuroimaging Initiative. The residual center of mass: an image descriptor for the diagnosis of Alzheimer disease. *Neuroinformatics*. 2019;17:307–321.
 33. Katoko A, Shelton P, Goertzen AL, et al. Machine learning identified an Alzheimer's disease-related FDG-PET pattern which is also expressed in Lewy body dementia and Parkinson's disease dementia. *Sci Rep*. 2018;8:13236.
 34. Liu M, Cheng D, Yan W; Alzheimer's Disease Neuroimaging Initiative. Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Front Neuroinform*. 2018;12:35.
 35. Kim J, Lee B. Identification of Alzheimer's disease and mild cognitive impairment using multimodal sparse hierarchical extreme learning machine. *Hum Brain Mapp*. May 7, 2018 [Epub ahead of print].
 36. Lu D, Popuri K, Ding GW, Balachandrar R, Beg MF; Alzheimer's Disease Neuroimaging Initiative. Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. *Sci Rep*. 2018;8:5697.
 37. Liu M, Cheng D, Wang K, Wang Y; Alzheimer's Disease Neuroimaging Initiative. Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. *Neuroinformatics*. 2018;16:295–308.
 38. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. arXiv.org website. <https://arxiv.org/abs/1512.00567>. Accessed June 20, 2019.
 39. Ding Y, Sohn JH, Kawczynski MG, et al. A deep learning model to predict a diagnosis of Alzheimer disease by using ¹⁸F-FDG PET of the brain. *Radiology*. 2019;290:456–464.
 40. 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: 887–893.
 41. Li S, Jiang H, Wang Z, Zhang G, Yao Y-D. An effective computer aided diagnosis model for pancreas cancer on PET/CT images. *Comput Methods Programs Biomed*. 2018;165:205–214.
 42. Perk T, Bradshaw T, Chen S, et al. Automated classification of benign and malignant lesions in ¹⁸F-NaF PET/CT images using machine learning. *Phys Med Biol*. 2018;63:225019.
 43. Roccaia E, Mikhno A, Ogden T, et al. Quantifying brain [¹⁸F]FDG uptake non-invasively by combining medical health records and dynamic PET imaging data. *IEEE J Biomed Health Inform*. 1, January 2019 [Epub ahead of print].
 44. Zhao L, Lu Z, Jiang J, Zhou Y, Wu Y, Feng Q. Automatic nasopharyngeal carcinoma segmentation using fully convolutional networks with auxiliary paths on dual-modality PET-CT images. *J Digit Imaging*. 2019;32:462–470.
 45. Huang B, Chen Z, Wu P-M, et al. Fully automated delineation of gross tumor volume for head and neck cancer on PET-CT using deep learning: a dual-center study. *Contrast Media Mol Imaging*. 2018;2018:8923028.
 46. Zhao X, Li L, Lu W, Tan S. Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. *Phys Med Biol*. 2018;64:015011.
 47. Zhong Z, Kim Y, Plichta K, et al. Simultaneous cosegmentation of tumors in PET-CT images using deep fully convolutional networks. *Med Phys*. 2019;46: 619–633.
 48. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology*. 2018;286:887–896.
 49. Do BH, Langlotz C, Beaulieu CF. Bone tumor diagnosis using a naïve Bayesian model of demographic and radiographic features. *J Digit Imaging*. 2017;30:640–647.
 50. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging*. 2016;35:1207–1216.
 51. Baessler B, Luecke C, Lurz J, et al. Cardiac MRI texture analysis of T1 and T2 maps in patients with infarctlike acute myocarditis. *Radiology*. 2018;289:357–365.
 52. Togo R, Hirata K, Manabe O, et al. Cardiac sarcoidosis classification with deep convolutional neural network-based features using polar maps. *Comput Biol Med*. 2019; 104:81–86.
 53. Ma L, Ma C, Liu Y, Wang X. Thyroid diagnosis from SPECT images using convolutional neural network with optimization. *Comput Intell Neurosci*. 2019;2019: 6212759.
 54. Ellmann S, Seyler L, Evers J, et al. Prediction of early metastatic disease in experimental breast cancer bone metastasis by combining PET/CT and MRI parameters to a model-averaged neural network. *Bone*. 2019;120:254–261.
 55. Chen L, Zhou Z, Sher D, et al. Combining many-objective radiomics and 3D convolutional neural network through evidential reasoning to predict lymph node metastasis in head and neck cancer. *Phys Med Biol*. 2019;64:075011.
 56. Xiong J, Yu W, Ma J, Ren Y, Fu X, Zhao J. The role of PET-based radiomic features in predicting local control of esophageal cancer treated with concurrent chemoradiotherapy. *Sci Rep*. 2018;8:9902.

57. Milgrom SA, Elhalawani H, Lee J, et al. A PET radiomics model to predict refractory mediastinal Hodgkin lymphoma. *Sci Rep*. 2019;9:1322.
58. Peeken JC, Goldberg T, Pyka T, et al. Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme. *Cancer Med*. 2019;8:128–136.
59. Papp L, Pötsch N, Grahovac M, et al. Glioma survival prediction with combined analysis of in vivo ¹¹C-MET PET features, ex vivo features, and patient features by supervised machine learning. *J Nucl Med*. 2018;59:892–899.
60. Ingrisich M, Schöppe F, Paprotka K, et al. Prediction of ⁹⁰Y radioembolization outcome from pretherapeutic factors with random survival forests. *J Nucl Med*. 2018;59:769–773.
61. Betancur J, Hu LH, Commandeur F, et al. Deep learning analysis of upright-supine high-efficiency SPECT myocardial perfusion imaging for prediction of obstructive coronary artery disease: a multicenter study. *J Nucl Med*. 2019;60:664–670.
62. Betancur J, Otaki Y, Motwani M, et al. Prognostic value of combined clinical and myocardial perfusion imaging data using machine learning. *JACC Cardiovasc Imaging*. 2018;11:1000–1009.
63. Kumar A, Kim J, Cai W, Fulham M, Feng D. Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data. *J Digit Imaging*. 2013;26:1025–1039.
64. Pelka O, Nensa F, Friedrich CM. Annotation of enhanced radiographs for medical image retrieval with deep convolutional neural networks. *PLoS One*. 2018;13:e0206229.
65. Sellam T, Lin K, Huang IY, et al. DeepBase: deep inspection of neural networks. arXiv.org website. <https://arxiv.org/abs/1808.04486>. Accessed June 20, 2019.
66. Gunning D. Explainable artificial intelligence (XAI). <https://www.darpa.mil/program/explainable-artificial-intelligence>. Accessed June 20, 2019.
67. Knight W. Google and others are building AI systems that doubt themselves. <https://www.technologyreview.com/s/609762/google-and-others-are-building-ai-systems-that-doubt-themselves/>. Accessed June 20, 2019.
68. Dillon J, Shwe M, Tran D. Introducing TensorFlow probability. <https://medium.com/tensorflow/introducing-tensorflow-probability-dca4c304e245>. Accessed June 20, 2019.
69. Robinson R, Valindria VV, Bai W, et al. Automated quality control in image segmentation: application to the UK Biobank cardiovascular magnetic resonance imaging study. *J Cardiovasc Magn Reson*. 2019;21:18.
70. Chapman EN, Kaatz A, Carnes M. Physicians and implicit bias: how doctors may unwittingly perpetuate health care disparities. *J Gen Intern Med*. 2013;28:1504–1510.
71. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. <http://science.sciencemag.org/content/356/6334/183>. Accessed June 20, 2019.
72. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*. 2018;81:77–91.
73. Hinton G. On radiology. <https://youtu.be/2HMpRXstSvQ>. Accessed June 20, 2019.
74. Davenport TH, Dreyer KJ. AI will change radiology, but it won't replace radiologists. *Harv Bus Rev*. March 27, 2018. <https://hbr.org/2018/03/ai-will-change-radiology-but-it-wont-replace-radiologists>. Accessed June 20, 2019.
75. Images aren't everything: AI, radiology and the future of work. *The Economist*. June 7, 2018. <https://www.economist.com/leaders/2018/06/07/ai-radiology-and-the-future-of-work>. Accessed June 20, 2019.
76. Parker W. Despite AI, the radiologist is here to stay. <https://medium.com/authorized-storytelling/the-radiologist-is-here-to-stay-24da650621b5>. Accessed June 20, 2019.
77. Chan S, Siegel EL. Will machine learning end the viability of radiology as a thriving medical specialty? *Br J Radiol*. 2019;92:20180416.
78. AI and future jobs: estimates of employment for 2030. https://techcastglobal.com/techcast-publication/ai-and-future-jobs/?p_id=11. Accessed June 20, 2019.
79. Will a robot take your job? <https://www.bbc.com/news/technology-34066941>. Accessed June 20, 2019.
80. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. *JAMA*. 2016;316:2353–2354.



The Journal of
NUCLEAR MEDICINE

Artificial Intelligence in Nuclear Medicine

Felix Nensa, Aydin Demircioglu and Christoph Rischpler

J Nucl Med. 2019;60:29S-37S.

Doi: 10.2967/jnumed.118.220590

This article and updated information are available at:
http://jnm.snmjournals.org/content/60/Supplement_2/29S

Information about reproducing figures, tables, or other portions of this article can be found online at:
<http://jnm.snmjournals.org/site/misc/permission.xhtml>

Information about subscriptions to JNM can be found at:
<http://jnm.snmjournals.org/site/subscriptions/online.xhtml>

The Journal of Nuclear Medicine is published monthly.
SNMMI | Society of Nuclear Medicine and Molecular Imaging
1850 Samuel Morse Drive, Reston, VA 20190.
(Print ISSN: 0161-5505, Online ISSN: 2159-662X)

© Copyright 2019 SNMMI; all rights reserved.