Artificial Intelligence in the Analysis of PET Scans of the Human Brain

5

Kim Mouridsen and Ronald Borra

Contents

5.1 Introduction 106
5.2 Artificial Intelligence 107
5.3 Scope of Classical Statistics Versus Deep Learning 109
5.4 Inferring CT and MR Information 110
5.5 Image Reconstruction 111
5.6 Dose Reduction 112
5.7 Automated Detection of Pathology 112
5.8 Considerations for the Future 114
References 115

Abstract

Artificial intelligence (AI) is being intensely studied, evaluated, and applied in healthcare and especially in medical imaging. Having shown performance equaling that of experienced radiologists in tasks such as detecting pneumonia on chest X-rays and identifying cancerous long nodules on X-ray, AI is poised to radically optimize many areas of medical practice from early detection of disease to prediction of progression and personalization of therapeutic strategy. Artificial intelligence extends classical statistical techniques and machine learning, both of which characteristically involve manually establishing imaging features hypothesized to modulate a certain outcome. With AI, predictive features are automatically established in a data-driven fashion, which in turn implies that raw

K. Mouridsen (✉)
Center of Functionally Integrative Neuroscience and MINDLab, Inst. of Clinical Medicine, Aarhus University, Aarhus C, Denmark
e-mail: kim@cfin.au.dk

R. Borra
University Medical Center Groningen, Groningen, The Netherlands

R. A. J. O. Dierckx et al. (eds.), *PET and SPECT in Neurology*,
https://doi.org/10.1007/978-3-030-53168-3_5

unprocessed data can be fully utilized, human bias can be avoided, and previously unrealized disease mechanisms potentially can be discovered. Here we discuss applications of AI in PET imaging for image reconstruction, attenuation correction without CT, dose reduction, automatic identification of pathology, and differentiation of disease progression. One of the costs of more automated analyses and better accuracy with AI compared to classical machine learning is larger volumes of training data; however, the field is rapidly evolving, and we discuss possible mitigations as well as other directions for valuable future applications of AI in PET imaging.

5.1 Introduction

Deep learning may be considered the big bang moment for artificial intelligence (AI). The idea of mimicking human brain architecture to build "intelligent" machines dates back at least to the period 1930–1950 where research in neurology showed that the brain consists of networks of neurons exchanging electrical signals and concurrently researchers showed that any form of computation can be performed digitally. Over the years, however, mathematical models based on idealized neurons organized in a hierarchical structure were realized to be unstable, brittle, in the sense that training the same network more than once led to different solutions and networks could yield surprising results given previously unseen input.

The pursuit for electronic brains gave way for less ambitious or "flat" models such as support vector machines and even classical logistic regression for tasks where a set of input variables should be combined to form a classification or prediction. In general interest in neural networks was replaced by developments in what is often referred to as classical machine learning. Despite a much simpler structure, classical machine learning techniques often demonstrate good classification or predictive accuracy, have been extensively applied in medical imaging, and lie at the heart of the relatively new field of radiomics, where tissue imaging features are manually crafted, for instance, with texture analysis, and combined into estimates of tissue type or presence of disease, for instance.

The idea that superior performance could be obtained by replicating neuronal architecture was not entirely abandoned, but it was not until the period 2006–2012 that neural networks accomplished the breakthroughs leading to today's enthusiasm. In this period Geoffrey Hinton, among others, discovered robust methods to train neural networks and laid the foundation for the field now known as deep learning. It centers on artificial neural networks with "neurons" or nodes, as they are called, organized in hierarchically connected layers, of which there can be hundreds and the total number of free parameters on the scale of millions; see Figure 5.1b for a schematic illustration.

As an indicator of progress, the annual ImageNet challenge considers a database of one million images from one thousand categories. By 2011 only classical machine learning had been attempted with the best algorithms reaching an accuracy around 75%. In 2012 deep learning yielded a remarkable 84.3% accuracy (Krizhevsky et al.

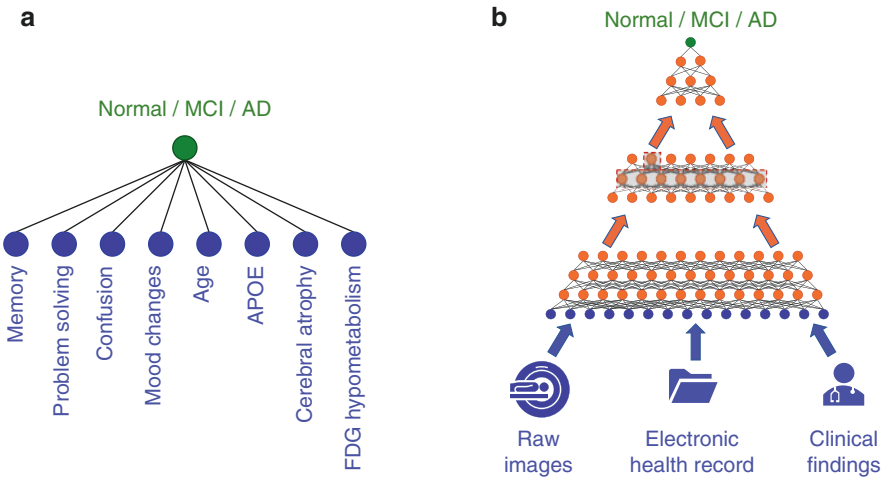

Fig. 5.1 (**a**) Graphic representation of the idea of combining a number of input variables into a single value. (**b**) Schematic illustration of an artificial neural network

2012). Perhaps the most surprising development was in 2015 where a deep neural network was reported to exceed human performance in image classification (He et al. 2015).

Clearly a natural domain for application of deep learning is medical imaging, and the technology has already demonstrated notable success. One study showed that a deep neural network (DNN) outperformed four radiologists in detecting pneumonia on chest X-rays (Rajpurkar et al. 2017). Another study found that a DNN outperformed 17 out of 18 radiologists in a task of identifying cancerous long nodules on X-rays (Hwang et al. 2019). With increasing demand for value and efficiency in healthcare, and not least medical imaging, AI is poised—inter alia—to improve early disease detection, accelerate image reading, and guide therapeutic decision-making.

As a crucial component in routine diagnostics as well as clinical trials, nuclear medicine stands to make important advances in combination with AI methodology. So far the field appears less explored than other advanced imaging modalities. A search on PubMed (October 18, 2019) for deep learning revealed only 65 publications with positron emission tomography against 271 with computed tomography and 420 with magnetic resonance imaging. In the following, we discuss current research directions, results, and future opportunities.

5.2 Artificial Intelligence

To better understand the potential and opportunities for AI in PET imaging, we offer a brief introduction and overview of methods from a "user's" perspective. Although the terminology is ambiguous, artificial intelligence is often used as an overarching term for methods aimed at identifying and parameterizing patterns in data. This can

be as simple as a classical regression model, where a number of observed quantities are combined to form an estimate or prediction of a target quantity. In a linear regression, this target is a value such as the volume of a tissue type or the level of metabolism. If the target is a category, such as presence or absence of disease, the same linear combination can be formed but is then transformed to represent the probability for each of the two, or more, classes. This is called logistic regression. Figure 5.1a displays graphically the idea of combining a number of input variables into a single value.

As we will see below, the state-of-the-art deep learning models, which can be said to be at the other end of the spectrum of mathematical complexity compared to regression, consist fundamentally of exactly the same simple building blocks, namely, linear combinations of predictors. This is illustrated in Figure 5.1b, where the small highlighted part of the larger network has exactly the structure of a regression model shown in Figure 5.1a. The difference is that deep neural networks comprise hierarchies of non-linear transformations of the basic linear combinations.

One advantage is that throughout the layers the model itself is able to construct combinations of input variables which are predictive for the target outcome. This allows tremendous flexibility in transforming—or encoding—the connection between predictors and response, where these hierarchical models also have the capacity to naturally uncover complex interactions across predictors. Interactions are also possible in classical regression models, but they must be specified manually a priori, and typically only subsets of pairs of variables are considered practically feasible.

Another advantage is that since it is possible to have even very high numbers of nodes in the input layer, it is feasible to input raw data such as image volumes (PET, MRI, CT), text (electronic health records, clinical findings), and signals (EEG, MEG). In contrast, with classical regression, raw data is typically filtered, aggregated, and subjected to manual feature extraction until rather few representative quantities are left. Broadly speaking, a regression model can in this case be seen as the result of human feature engineering followed by statistical fitting, whereas a neural network automates the entire process. The manual steps in classical data analyses prompt risk of bias, missing critical information and even overfitting, compromising generalizability. The latter is an often overlooked problem in classical scientific data analyses. Whereas the regression model itself is thoroughly treated by statistical software and estimates are properly corrected for level of data noise and number of parameters, the many choices of which parameters to include and how they are transformed and combined are not factored into the final results, although substantially influencing the degrees of freedom and potentially limiting reproducibility (Simmons et al. 2011; Ioannidis 2005).

A third advantage is that deep learning models have the capability to continue improving in performance as the volume of input data increases. Simple models such as regressions are naturally limited in their ability to model complex relations between inputs and outputs, whereas deep neural networks, with up to millions of parameters, are able to encode much more complex relations, such as the

differences in disease progression across diverse populations. The volume of parameters is at the same time the major limitation of deep neural networks when data is scarce, where classical models may be more robustly applied.

5.3 Scope of Classical Statistics Versus Deep Learning

As we have seen, modern deep neural networks extend in terms of architecture the classical regression model. However, the purpose and the methods for fitting the two approaches differ considerably. From a traditional statistical point of view, linear combinations are used in an inferential way to quantify the degree (and type) of influence a predictor seemingly exerts on an outcome measure. This has been qualified, for instance, by the amount of change in response that would be expected by a one-unit change in the value of the predictor. Formally, significance tests have been used as an indicator for whether such an effect should be considered manifest or could have occurred by chance.

Considering regression as a (very) simple machine learning model means changing one's perspective. Instead of focusing on quantification of individual predictor's effect on the response, we focus on making the most accurate predictions. For this purpose, the inferential question of whether a predictor is statistically significant is less relevant, and instead the focus is on the collective predictive ability. This could be conceived as a loss, but if, for instance, the input for a prediction, e.g., prediction of progression of dementia, is an entire image, it can be argued that it is not the individual voxels but rather their combined effects and interactions which are of interest. In this case the notion of effect of a single element may be less meaningful in that the voxel may have negligible influence in itself but may contribute via complex interactions with other tissue regions. The matter of whether an element provides actual information, however, persists but manifests in a different way. Namely, if a predictor does not consistently contribute to prediction of the response, its coefficient parameter will typically be low or zero. This is ensured by a mechanism during model development referred to as cross-validation, which may be seen as an even stricter control than traditional statistical inference based on *p*-values. During model development data used for fitting is typically split in a training partition and a validation partition. This systematically simulates the real-world setting where the model must make predictions for new and previously unseen cases. Hence if a predictor has an apparently strong effect on the outcome in the training data, but performance on the validation data is low, then the estimated strength of each predictor is adjusted accordingly. As a result, the final model will have "proven" that it performs well on the data on which it was trained but is also expected to perform well on unseen cases. This is a very important and valuable criteria especially for methods intended to be used in actual clinical settings. It is important to note that in traditional statistics, all data points are typically used for model fitting and *p*-values used for inference are based on in-sample estimates. Hence it can be argued that artificial intelligence algorithms undergo a stricter scrutiny than traditional

statistical models and foster greater requirements for predictors than traditional inference.

As an additional note, *p*-values are influenced not only by the size of the effect they are associated with but also by the number of observations in the study. Accordingly, as studies increase in volume of samples, assuming that the effect of a predictor remains constant, the *p*-value will in any case become smaller. Therefore even the most scientifically negligible effects will eventually become statistically significant. In contrast, the actual influence and therefore predictive ability of the variable remain the same, and therefore the performance of the model will not change (Mouridsen 2015). Measurement of model performance can therefore be argued to give a more robust picture of the face value of predictors.

5.4 Inferring CT and MR Information

A challenge for quantitative PET imaging is the spatially accurate source of photon emission. The tissue-dependent attenuation, such as Compton scatter, of the gamma-rays compromises precise volumetric reconstruction. A common resolution to this problem is to acquire an additional CT, typically in combined PET/CT systems, which yield clear identification of bone structures which can be used as a basis to establish a photon attenuation correction map. However this may not always be a practical solution wherefore deep learning has been considered to alleviate this problem in different settings.

For instance, with combined PET/MR scanners, a CT scan for attenuation correction is unavailable. At the same time, while standard MRI sequences are excellent at displaying soft-tissue contrast, MRI poorly displays bone and can therefore not be used per se as a basis for attenuation correction. This can be mediated to some extent using very short echo times, but errors remain in the attenuation correction. As an alternative it has been suggested to generate a pseudo-CT image from MRI (Liu et al. 2018a). With 40 subjects for whom T1 MRI as well as plain CT was acquired, an encoder-decoder type neural network was constructed aiming to input an MR image and output a segmented CT image. The pseudo-CT images were shown to match actual segmented CT with a Dice coefficient of 0.803 ± 0.021 for bone. In a prospective study with five patients, a significantly lower PET reconstruction error was demonstrated with deep learning-based attenuation correction than with standard Dixon-based soft-tissue and air segmentation and anatomic CT-based template registration.

An alternative idea is to directly estimate a pseudo-CT image with tissue segmentation from the uncorrected PET image, thereby avoiding the extra step involved in acquiring a second image modality. One procedure for this was recently proposed (Liu et al. 2018b). This study uses a similar encoder-decoder structure network as in Liu et al. (2018a) except for the addition of short-cut connections making the architecture appear comparable to the U-net (Ronneberger et al. 2015). The network was trained based on 100 subjects and validated on an additional 28. The authors report a Dice coefficient for bone segmentation of 0.75 ± 0.03 compared to actual non-contrast CT and ^{18}F-FDG-PET errors less than 2% compared to CT.

Looking beyond attenuation correction, the possibility for image-to-image translation may also be useful, for instance, in diagnostic workup where two modalities are necessary. For instance, amyloid deposition is a useful biomarker in patients suspected of cognitive decline, where absence of amyloid depositions in the cortex can rule out Alzheimer's disease. Being established as SUV relative to a reference region, PET imaging should therefore ideally be supplemented by a structural MR acquisition. However, this may not be feasible in the clinic. As an alternative (Choi and Lee 2018) suggests producing a pseudo-T1 MRI from ^{18}F-florbetapir PET. The model was trained using 163 patients from the ADNI study. The authors used a so-called generative adversarial neural network which is an approach in which two competing networks are trained in parallel (Goodfellow et al. 2014). One is trained with the aim to generate realistic MR images from PET, while the other is trained to determine whether an MR image is artificially generated or actually acquired**.** Performance test in 98 subjects from 8 independent sites showed an absolute error in standardized SUV significantly smaller than for other MR-less approaches such as template matching and segmentation.

5.5 Image Reconstruction

PET images are created via a complex process where scintillation in the inorganic crystals in the PET scanner is converted to spatially resolved maps of metabolic activity. This process holds a number of possible applications for artificial intelligence. Raw data produced in a PET acquisition are in the form of a sinogram representing the number of photon coincidence detections in pairs of detectors. Conversion from sinogram data to tissue maps is known as image reconstruction. This can be accomplished with so-called filtered backprojection (FBP) which is computationally efficient but has a range of limitations including tendency to produce streaks across images from high-uptake areas and sensitivity to shot noise. Statistical methods acknowledge the randomness in the underlying physical processes and avoid the artifacts of FBP but are slow because they require iterative maximization of a likelihood function, for instance, using expectation maximization (MLEM).

In Cui et al. (2017), the authors suggest improving the MLEM by using stacked autoencoders to automatically extract important image features. However it has also been suggested to entirely bypass traditional image reconstruction and simply construct metabolic maps from raw sinogram data (Häggström et al. 2019). By presenting a neural network on the input side with sinogram data and on the output side spatial maps established with traditional, but time-consuming, image reconstruction, the authors demonstrate that the network becomes capable of prospectively producing spatial maps from sinogram data with similar quality but more than 100 times faster.

PET image quality depends on the features of the scintillating crystals. More coarsely pixelated crystals are less sensitive to noise but produce blurred images, whereas thin-pixelated crystals yield higher spatial accuracy but at the cost of noise sensitivity. Hong et al. (2018) suggest that comparable image resolution and better

noise properties can be obtained with large crystals of bin sizes a factor 4 larger than thin crystals using a deep residual convolutional neural network. For an application of denoising with combined PET/CT data, see, for instance, Cui et al. (2019).

5.6 Dose Reduction

One of the main limitations for wider use of PET in the clinic is the necessity of a radioactive tracer. Due to radiation exposure, PET imaging is often limited to terminally ill or elderly patients. If the radiation dose could be reduced without substantially compromising image quality, the more general clinical value of PET could increase tremendously.

With image quality depending on the balance of bed duration and radiation dose, early work (Xiang et al. 2017) used a reduction in acquisition time from 12 min to 3 min following a standard dose of ^{18}F-2-deoxyglucose to simulate the effect of administering a dose 75% below the standard. Using a concatenation of convolutional neural networks with the pseudo-low-dose (i.e., short scan duration) images as input and standard-dose (long scan duration) images as output to predict, the study shows in 16 patients that in particular when the low-dose PET is complemented by T1 MR competitive image quality is seen relative to a state-of-the-art technique however 500 times faster.

Later, Xu et al. (2017) administered standard-dose ^{18}F-fluorodeoxyglucose (370 MBq) in nine patients with glioblastoma with a PET/MR scanner. Low-dose PET was simulated by randomly selecting only 0.5% of the count events based on raw listmode data corresponding to a 200× reduction in dose. The authors used so-called residual learning where a neural network is trained to learn the difference between low-dose PET as input and standard-dose PET as output. This is found to be a more robust training strategy for deep networks. The resulting network demonstrated to produce superior performance in estimating standard-dose PET from low-dose PET compared to state-of-the-art techniques.

Clearly synthetic images reconstructed from lower resolution will not match standard acquisitions numerically exact, but for clinical purposes, a more relevant question is whether similar diagnostic conclusions are reached with the synthetic images. To study this, Chen et al. (2019) considered amyloid (fluorine 18 [^{18}F]-florbetaben) PET/MR acquisitions in 39 subjects and a neural network architecture similar to Xu et al. (2017). Expert human assessment of amyloid status with synthetic images yielded good accuracy compared to full-dose images with an accuracy of 89% corresponding to 71 out of 80 readings and was found to be similar to intrareader reproducibility of full-dose images (73 of 80 [91%]).

5.7 Automated Detection of Pathology

Combined PET and CT imaging plays a major role in managing oncological patients and in particular for radiation therapy planning. Outlining the gross tumor as well as pathological lymph nodes is a time-consuming manual task. In addition to the

time requirement, the manual intervention also limits reproducibility and standardization. Automation of this task is therefore potentially a clinically valuable application of artificial intelligence. This task has previously been approached with traditional machine learning methods such as k-nearest neighbors, Markov random fields, adaptive random walk with k-means, and decision trees (Yu et al. 2009; Comelli et al. 2018; Yang et al. 2015; Stefano et al. 2017; Berthon et al. 2017).

Using a deep learning approach, Moe et al. (2019) describe a method for automatic delineation of gross tumor volume as well as pathological lymph nodes. This is based on the often used U-Net convolutional architecture (Ronneberger et al. 2015)**.** Imaging regimen included FDG-PET and CT. The best correspondence with manual outlining was seen when using PET and CT images in combination, followed by PET only. The Dice coefficient for PET/CT was 0.75 ± 0.12, whereas for CT only this dropped to 0.65 ± 0.17. This was based on a training set of 142, validation set of 15, and test set of 42 patients stratified on tumor T-stage.

In a smaller study (Huang et al. 2018), however with combined PET and CT data from two sites, a convolutional architecture also based on the U-net was proposed to segment gross tumor volume. To mitigate the challenge of a lower number of patients, this study explicitly used simulated data augmentation in the form of rotating the images, horizontal mirroring, changing the contrast, and image scaling. The resulting median Dice score as calculated with leave-one-out cross-validation was 0.785 with a range of 0.482 to 0.868.

Correct delineation of tumor volume in clinical practice also depends on the ability to differentiate between lesions due to tumor progression and tissue damage due to radiotherapy. For instance, radiosurgery treatment of patients with brain metastases may result in radiation necrosis rates between 25% and 50% (Minniti et al. 2011). Mis-interpretation of treatment-induced change as tumor progression may lead to unwarranted treatment approaches or cessation of effective therapy as well as biased estimation of therapeutic success in clinical trials.

Whereas reports (Chao et al. 2001) on the use of FDG-PET for differentiating metastasis relapse from radiation damage are ambiguous on the clinical value, with sensitivities and specificities ranging from 40% to 100%, the MET and FDOPA PET have more consistently demonstrated clinical feasibility with sensitivities and specificities around 80%. Similarly for the differentiation of treatment-related damage and progression of glioma, PET imaging with FET and FDOPA has shown clinically relevant performance. While artificial intelligence may be hypothesized to improve these performances by a more detailed analysis of lesion extent and texture notably the implementation of dynamic PET (Galldiks et al. 2012), where subtle patterns in the uptake curves may hold further clues, which can be exploited by computationally intensive methods.

Detection of disease can also have a more global aim than regional gross tumor or metastasis identification and delineation. For instance, detection of Alzheimer's disease (AD) is important for diagnostic purposes and is expected to become even more critical with the advent of therapies to potentially delay onset, slow progression, or even cure the disease. At the heart of this task is separation of prodromal AD cases from patients with mild cognitive impairment (MCI) who will not progress to AD. It is speculated that a deeper understanding of the progression from MCI to AD

is necessary also at the time of development of pharmacological interventions—since only a fraction of patients with symptoms of MCI develop AD, the patients who would not progress to AD will seemingly dilute actual therapeutic effects for those who will progress. Being able to design studies such that only patients predicted to progress to AD from MCI are included will increase the statistical power for detection of actual effects.

FDG-PET plays an important role in the diagnosis of AD due to its ability to show metabolic activity which is recognized to preceed anatomical changes, as observed with MRI, and account for the cognitive and functional decline observed with the disease. However, success in developing automated methods for predicting which MCI cases will progress to AD has been limited with accuracies typically not exceeding 80% (Lu et al. 2018).

One larger study (Lu et al. 2018) considers 1.051 patients from the publicly available ADNI dataset with structural MRI as well as FDG-PET. Subjects were divided in normal controls, stable mild cognitive impairment (sMCI) class, progressive mild cognitive impairment (pMCI) class, and those clinically diagnosed with Alzheimer's disease. A deep learning approach was suggested which simultaneously analyzes images across different spatial resolutions. This produces a number of classifiers, which are then combined into a single result via so-called ensemble learning. An accuracy of 82.51(5.35) was demonstrated for automatically differentiating between stable and progressive MCI. For the classification of NC versus AD, the accuracy reported was 93.58 (5.20). The principle of combining multiple classifications into one has also previously been explored in a subset of the ADNI data (Ortiz et al. 2016).

This study employed a strategy of pre-training the neural network using a so-called autoencoder approach. Autoencoders have previously shown efficiency in detection of AD in a smaller subset of 311 subjects from the ADNI data (Liu et al. 2014). One of the challenges in deep learning models, which contain a large number of unknown parameters, is finding suitable starting values (known as initialization). An autoencoder can be used to mitigate this problem by effectively reducing the complex information in raw data to more compact distillation. This is accomplished by channeling the high-dimensional raw data through a lower-dimensional layer of hidden units in neural network and then optimizing this layer such that it can at the same time reproduce the input data. This is similar to traditional principal component analysis, where raw data is reduced to principal components which are linear combinations of data input. Autoencoders generalize this approach by using non-linear transformations. The many layers in a deep neural network can now be initialized in a bottom-up approach where each layer is initialized by the hidden layer of an appropriate autoencoder.

5.8 Considerations for the Future

Across medical imaging and healthcare in general, there are a seemingly endless stream of new ideas for application of AI and a corresponding multitude of proof-of-concept-level studies. However the uptake in clinical practice is still minimal. In 2017 and 2018, only around 14 algorithms were approved by the Food and Drug Administration (FDA) in the USA. It is a concern that few of the companies behind

these products have published peer-reviewed papers let alone conducted prospective clinical trials. This shows a trend where validation of technology as well as actual clinical value is lacking.

One reason for the lack of validation studies is that datasets in medical imaging are typically small. While some of the largest medical imaging datasets used for deep learning contains in the order of 100.000 patients, run-of-the-mill research studies are much smaller with cohort sizes from around 30 to a few thousands. This may be due to limited funding, but privacy concerns also make compilation of large data volumes from different sites across the world let alone commercial participation difficult. Research publications, such as those reviewed here, do offer validations, but only the so-called cross-validations which are essentially made by randomly splitting the dataset into training and testing. The testing dataset therefore, statistically, has the same properties as the training data which may not be the case if new data is acquired with another machine or at another hospital. Also inclusion and exclusion criteria must be exactly identical. For clinical uptake not only larger datasets with representative samples are critical to show consistent high accuracy, but the technologies must be tested prospectively in the clinic to show actual value. With smaller datasets the risk of bias also increases, and bias in AI is a concern. However in fairness this should be compared to human bias which has also been reported to be significant (Topol 2019). In the case of AI, the bias is in principle avoidable by having representative data samples. Finally although cross-validation is the typical performance metric used in AI studies, the field lacks a standard for how this is performed. Without such standardization, it is difficult to compare performance across studies. In classical statistics, *p*-values have the role of a standardized measure of significance. Although such single measures may be misused or overinterpreted, as has been the case with *p*-values (Ioannidis 2005), efforts toward developing standardized reporting practices will be beneficial.

For the first wave of AI enthusiasm algorithm performance has been the center of attention. To move the field forward, increase trust in AI technology, and increase clinical acceptance, the pursuit to develop explainable AI will likely have a substantial role. A common criticism of deep neural networks is that they are black boxes unable to account for their otherwise high accuracy. In Europe, the General Data Protection Regulation (GDPR) stipulates that decisions relating to a patient may not fully rely on automated processing, often taken to mean that patients have a right to an explanation for decisions. As the field moves forward, the ability of algorithms to account for their decisions may not only benefit patients but also be the source of many new insights and discoveries based on combinations of imaging and health record data made possible by modern versions of artificial intelligence.

References

Berthon B et al (2017) Head and neck target delineation using a novel PET automatic segmentation algorithm. Radiother Oncol 122:242–247

Chao ST, Suh JH, Raja S, Lee SY, Barnett G (2001) The sensitivity and specificity of FDG PET in distinguishing recurrent brain tumor from radionecrosis in patients treated with stereotactic radiosurgery. Int J Cancer 96:191–197

Chen KT et al (2019) Ultra-low-dose (18)F-Florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. Radiology 290:649–656
Choi H, Lee DS (2018) Generation of structural MR images from amyloid PET: application to MR-less quantification. J Nucl Med 59:1111–1117
Comelli A, Stefano A, Benfante V, Russo G (2018) Normal and abnormal tissue classification in positron emission tomography oncological studies. Pattern Recogn Image Anal 28:106–113
Cui J, Liu X, Wang Y, Liu H (2017) Deep reconstruction model for dynamic PET images. PLoS One 12:e0184667
Cui J et al (2019) PET image denoising using unsupervised deep learning. Eur J Nucl Med Mol Imaging 46(13):2780–2789
Galldiks N et al (2012) Role of O-(2-(18)F-fluoroethyl)-L-tyrosine PET for differentiation of local recurrent brain metastasis from radiation necrosis. J Nucl Med 53:1367–1374
Goodfellow IJ et al (2014) Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, vol 2. MIT Press, Montreal, pp 2672–2680
Häggström I, Schmidtlein CR, Campanella G, Fuchs TJ (2019) DeepPET: a deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. Med Image Anal 54:253–262
He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, Washington, D.C., pp 1026–1034
Hong X et al (2018) Enhancing the image quality via transferred deep residual learning of coarse PET Sinograms. IEEE Trans Med Imaging 37:2322
Huang B et al (2018) 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 2018:8923028
Hwang EJ et al (2019) Development and validation of a deep learning-based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs. Clin Infect Dis 69:739–747
Ioannidis JP (2005) Why most published research findings are false. PLoS Med 2:e124
Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, vol 1. Curran Associates Inc., Lake Tahoe, Nevada, pp 1097–1105
Liu S et al (2014) 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI). IEEE, Piscataway, pp 1015–1018
Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB (2018a) Deep learning MR imaging-based attenuation correction for PET/MR imaging. Radiology 286:676–684
Liu F et al (2018b) A deep learning approach for (18)F-FDG PET attenuation correction. EJNMMI Phys 5:24
Lu D, Popuri K, Ding GW, Balachandar R, Beg MF (2018) Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. Med Image Anal 46:26–34
Minniti G et al (2011) Stereotactic radiosurgery for brain metastases: analysis of outcome and risk of brain radionecrosis. Radiat Oncol 6:48
Moe Y et al. (2019) Deep learning for automatic tumour segmentation in PET/CT images of patients with head and neck cancers
Mouridsen K (2015) Behavioral methods in consciousness research. Oxford University Press, Oxford
Ortiz A, Munilla J, Gorriz JM, Ramirez J (2016) Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. Int J Neural Syst 26:1650025
Rajpurkar P et al. (2017), CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning
Ronneberger O, Fischer P, Brox T (2015) In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Springer International Publishing, Cham, pp 234–241

Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci 22:1359–1366

Stefano A et al (2017) An enhanced random walk algorithm for delineation of head and neck cancers in PET studies. Med Biol Eng Comput 55:897–908

Topol E (2019) Deep medicine: how artificial intelligence can make healthcare human again. Basic Books Inc., New York

Xiang L et al (2017) Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing 267:406–416

Xu J, Gong E, Pauly J, Zaharchuk G (2017) 200x Low-dose PET Reconstruction using Deep Learning

Yang J, Beadle BM, Garden AS, Schwartz DL, Aristophanous M (2015) A multimodality segmentation framework for automatic target delineation in head and neck radiotherapy. Med Phys 42:5310–5320

Yu H, Caldwell C, Mah K, Mozeg D (2009) Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. IEEE Trans Med Imaging 28:374–383