



Deep learning in Nuclear Medicine—focus on CNN-based approaches for PET/CT and PET/MR: where do we stand?

Margarita Kirienko¹ · Matteo Biroli² · Fabrizia Gelardi² · Ettore Seregni¹ · Arturo Chiti^{2,3} · Martina Sollini^{2,3}

Received: 14 October 2020 / Accepted: 8 January 2021 / Published online: 28 January 2021
© Italian Association of Nuclear Medicine and Molecular Imaging 2021

Abstract

Introduction In recent years, machine learning algorithms have led to innovative tools for medical imaging analysis. The purpose of the present review was to summarize the literature on the developing field of deep learning (DL), particularly the application of convolutional neural networks (CNNs) in PET/CT and PET/MR.

Methods We performed the literature search, referring to “convolutional neural networks” and “positron emission tomography” on PubMed/MEDLINE, for potentially relevant articles published up until July 24th, 2020.

Results After the screening process, 63 articles were finally included; these embraced both the technical ($n=23$) and the clinical field ($n=40$). Technical studies aimed at investigating the role of CNN-based methods for image quality improvement ($n=11$) and on technical issues ($n=12$), mainly attenuation correction. Clinical studies explored CNN applications in oncology lung cancer ($n=7$), head and neck cancer ($n=4$), esophageal cancer ($n=2$), lymphoma ($n=3$), prostate cancer ($N=4$), cervical cancer ($n=1$), sarcomas ($n=1$), multiple cancer types ($n=4$), in neurology ($n=10$) and cardiology ($n=1$); three additional studies belonged to “other” category. In oncology, the studies aimed at detection, diagnosis, and prognostication of cancer. In neurology, the majority of the studies aimed at diagnosing Alzheimer Disease and stratification of the risk. CNN-based algorithms demonstrated promising results with performances equal or even higher compared to conventional approaches.

Discussion Overall, CNN applications for PET/CT and PET/MR are exponentially growing, demonstrating encouraging results for both technical and clinical purposes. Novel research strategies emerged to face the challenges of DL algorithms development. Education and confidence with DL-based tools are needed for proper technology implementation.

Keywords Artificial intelligence · Machine learning · Convolutional neural networks · PET/CT · PET/MR

Purpose

Machine learning (ML) is a subfield of artificial intelligence (AI) primarily aimed at identifying patterns. Several ML algorithms can be applied, such as a support vector machine (SVM), decision trees, and Bayes network, but deep learning

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40336-021-00411-6>.

✉ Margarita Kirienko
margarita.kirienko@icloud.com

Matteo Biroli
matteo.biroli@st.hunimed.eu

Fabrizia Gelardi
fabrizia.gelardi@cancercenter.humanitas.it

Ettore Seregni
ettore.seregni@istitutotumori.mi.it

Arturo Chiti
arturo.chiti@hunimed.eu

Martina Sollini
martina.sollini@cancercenter.humanitas.it

¹ Nuclear Medicine, Fondazione IRCCS Istituto Nazionale Dei Tumori, Via Venezian 1, Milan, Italy

² Humanitas University, Pieve Emanuele, Milan, Italy

³ Humanitas Clinical and Research Center–IRCCS, Rozzano, MI, Italy

has achieved the most remarkable performance and success. In particular, for the image-based tasks, convolutional neural networks (CNN) dominate. This approach does not require handcrafted features calculation or operator input. The convolution operation uses multiple filters to extract features (feature map) from the input image. During training, CNNs learn those features that are critical to successful performance [1]. Figure 1 summarizes the main types of networks for imaging applications. Basic principles and definitions in ML are provided in the Supplementary material.

The use of CNNs has led to the development of image analysis algorithms for radiological applications: tuberculosis detection on chest X-ray, lung nodule and interstitial lung disease assessment on chest computed tomography (CT), pulmonary embolism identification on CT angiography, detection of breast mass on mammography, and of intracranial haemorrhage on head CT. Indeed, as of March 1st, 2020, radiology resulted in the most exploited field with the highest number of FDA approved tools based on ML technology [2].

Nuclear medicine is also expected to benefit from the CNN-based algorithms, particularly from tools for clinical decision support, examinations scheduling, proper imaging protocols choice, image quality improvement, interpretation and reporting. Therefore, the purpose of the present review was to summarize the available literature on the developing field of deep learning, particularly the application of CNNs, in PET/CT and PET/MR.

Materials and methods

Eligibility criteria, search strategy and study selection

Using the PubMed and MEDLINE database, we performed a comprehensive literature search for potentially relevant articles published up until July 24th, 2020. No limitations on publication date were applied. The search strategy combined terms (text words) referring to "convolutional neural networks" and "positron emission tomography. In particular, the following search strategy was applied: "convolutional neural network" OR "CNN" AND "positron emission tomography" OR "PET" OR "PET/CT" OR "PET/MR". Titles and abstracts of retrieved records have been screened independently by MK and AC. Exclusion criteria were: not-original articles, review articles, book chapters, editorials, case reports, non-English language papers, duplicates, non-in-human studies and studies out of the field of interest. Subsequently, we screened the reference list of selected studies to identify additional eligible articles. Aiming at a comprehensive assessment of the early stage of the development of deep learning applications in PET-scoping

review—no additional exclusion criteria, which assess quality, were used; consequently, we included early-stage and proof-of-concept investigations.

Data extraction and analysis

We summarized study characteristics for all the selected papers. Study characteristics included: title, authors, year of publication, abstract, study design, population (public dataset or not) and sample size, application (technical or clinical), medical field (oncology, neurology, cardiology or other) and disease/condition, type of imaging modality (PET/CT or PET/MRI), radiopharmaceutical, aim, and input type of data. According to the objective—technical vs diagnosis/prognostication—the articles were categorized into Image Quality and Technical applications vs Clinical studies. Main results and performance metrics were recorded. Descriptive statistical metrics were used to summarize the data.

Results

Study selection

The search of the PubMed/MEDLINE database returned a total of 381 studies. After the removal of duplicates, 110 records were left. After the abstract review and inclusion/exclusion criteria application, 47/110 studies were excluded. The screening process is summarized in Fig. 2. Sixty-three articles were finally included.

Study characteristics

The 63 included studies embraced both the technical ($n=23$) and the clinical field ($n=40$). Technical studies aimed at investigating the role of CNN-based methods focussing their attention on the image quality ($n=11$) and technical issues ($n=12$), mainly attenuation correction.

Clinical studies explored CNN applications in lung cancer ($n=7$), head and neck cancer ($n=4$), esophageal cancer ($n=2$), lymphoma ($n=3$), prostate cancer ($n=4$), cervical cancer ($n=1$), sarcomas ($n=1$), multiple cancer types ($n=4$), in neurology ($n=10$) and cardiology ($n=1$). Three clinical studies belonged to "other" category investigating CNN-based strategies in sex determination, cerebellum tracer uptake and in the improvement of cerebral blood flow measurement. Summary of the characteristics of the selected studies is provided in Table 1.

The input modalities were: PET ($n=29$, of which $n=1$ PET MIP and $n=3$ PET sinograms) PET and CT ($n=13$), CT ($n=6$), PET and MR ($n=4$), MR ($n=5$), floodmaps ($n=1$), Coincidence waveforms ($n=1$), MLAA-based

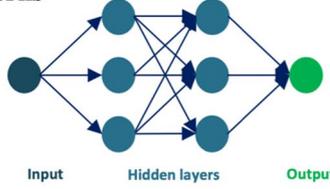
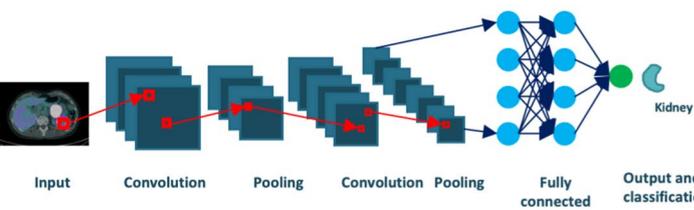
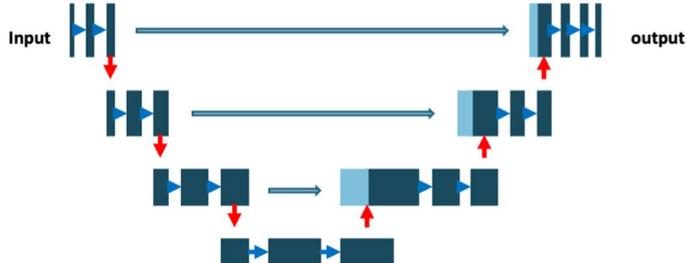
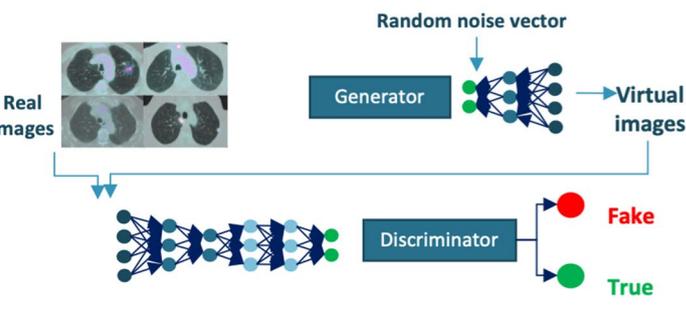
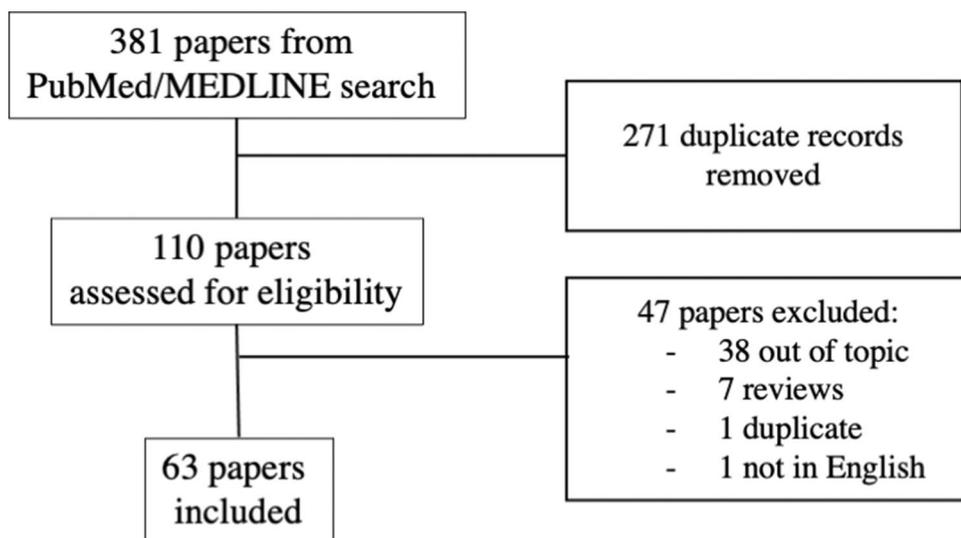
Neural Network – General Scheme	Description
<p>Artificial Neural Networks</p>  <p style="text-align: center;">Input Hidden layers Output</p>	<p>Intrinsic features of ANN are input, hidden and output layers. In Deep Neural Networks (DNN), each node, that performs a simple analysis, is connected to other nodes and constitute different layers. Each layer carries out calculations based on the output of the previous layer</p>
<p>Convolutional neural networks (CNNs)</p>  <p style="text-align: center;">Input Convolution Pooling Convolution Pooling Fully connected Output and classification</p>	<p>At the start, the series of convolution and pooling layers identify low-level high-resolution features; as the pools reduce the resolution, lower-level features are combined into higher-level features representing more complex structures. Fully connected layers are commonly used before the final, so that the high-level features are weighted to perform the task (e.g. classification or regression). Examples of CNNs are the original AlexNet, VGGNet, and GoogLeNet*.</p>
<p>U-Nets</p>  <p style="text-align: center;">Input output</p>	<p>U-Nets get their name from the appearance of the architecture scheme. At the bottom of the “U,” the image is reduced to the desired component, subsequently the refinement to original resolution is obtained with “bypass layers” (pixel data from the higher-resolution feature maps are used for upscaling) (e.g. SegNet belongs to U-Nets).</p>
<p>Generalized adversarial networks (GANs)</p>  <p style="text-align: center;">Random noise vector</p> <p style="text-align: center;">Generator Virtual images</p> <p style="text-align: center;">Real images Discriminator Fake True</p>	<p>Designed to create images rather than classify or segment them; this can be useful in medicine for the creation of additional training and testing images, and they may be useful for providing insight into how deep learning works, and for improving the robustness of deep learning systems. Two neural networks are trained in contest with each other. On the basis of a training set of images, the generator creates new virtual data, that are similar to training data. Simultaneously, the algorithm learns to distinguish fake data from true data.</p>
<p>* Refs for CNN examples: Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, CJC B, Bottou L, Weinberger KQ, editors. Advances in neural information processing systems 25. Red Hook, NY: Curran Associates; 2012. p. 1097–105. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. arXiv [cs.CV]. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Computer vision and pattern recognition (CVPR). 2015. http://arxiv.org/abs/1409.4842</p>	

Fig. 1 Main types of artificial neural networks for iamging applications

Fig. 2 Article selection process**Table 1** Summary of included studies' characteristics

Study characteristics	Technical Studies ($n=23$)	Clinical Studies ($n=40$)
Year of Publication		
Before 2017	$n=0$	$n=2$
2017	$n=2$	$n=2$
2018	$n=5$	$n=7$
2019	$n=9$	$n=18$
2020 (Jan–Jun)	$n=7$	$n=11$
Dataset size (patients/images)		
< 100	$n=13$	$n=16$
> 100 and < 500	$n=3$	$n=13$
> 500 and < 1000	$n=0$	$n=4$
> 1000	$n=0$	$n=5$
Not available	$n=7$	$n=2$
Study design		
Retrospective (Public datasets)	$n=11$ (1)	$n=32$ (9)
Prospective	$n=0$	$n=1$
Phantom	$n=5$	$n=0$
Both Phantom and Retrospective/ Public dataset	$n=2$	$n=0$
Not available	$n=5$	$n=7$
Imaging modality		
PET/CT	$n=15$	$n=36$
PET/MR	$n=7$	$n=4$
Not applicable	$n=1$	$n=0$

activity and attenuation maps ($n=2$), polar maps ($n=1$) and simulated PET-low-resolution sinogram ($n=1$).

Image quality and technical applications

Summary of the technical studies' main characteristics and findings is provided in Table 2.

Image quality

Radiation exposure is a central issue in nuclear medicine practice. A balance between the reduction of tracer activity and image quality is a challenge. On these grounds, Zhou et al. developed a supervised DL model (CycleGANs) to boost low-dose PET images quality. The proposed method was compared to other existing imaging denoising methods (Non-Local Mean (NLM) and block-matching 3D(BM3D), RED-CNN and 3D-cGAN). The proposed model accurately estimated full-dose PET image from low-dose input images, at the count level of 1 million true counts. Additionally, it preserved SUV_{mean} and SUV_{max} values and suppressed image noise for low dose PET imaging [3].

While Xiang et al. developed a CNN-based method to accurately estimate the standard PET image, combining both the low-quality low-dose PET (LPET) image and T1-weighted MR acquisition, the proposed method achieved a fast and competitive quality [4].

Spuhler et al. developed a denoising CNN-based method (dCNN) to recover full-count images from low-count images. dCNN was compared to existing conventional U-NET. The proposed algorithm achieved better results in terms of mean absolute percent error (MAPE): 4.99 ± 0.68 vs. 5.31 ± 0.76 ; peak signal-to-noise ratio (PSNR): 31.55 ± 1.31 dB vs. 31.05 ± 1.39 ; and structural similarity index metric (SSIM): 0.9513 ± 0.0154 vs. 0.9447 ± 0.0178 [5].

Image quality degradation and inaccurate image-based quantification related to the intrinsic PET low spatial

Table 2 CNN-based studies focussed on image quality and technical PET aspects

Research focus	Aim	Study Population	Study Design	Imaging modality (radiopharmaceutical)	CNN Input	Main results (mean ± SD or range)	References (publication year)
Image quality	Improve Low-Dose PET Imaging quality	18 patients	NA	PET/CT (¹⁸ F-FDG)	PET	NRMSE = 17.15 ± 3.27% SSIM = 0.30 ± 0.22% PSNR = 2.11 ± 1.67%	[3] (2020)
	Estimation of the high-quality standard-dose PET from low-dose PET and MR	16 patients	NA	PET/MR (¹⁸ F-FDG)	PET and MR	PSNR = 24.76 NMSE = 0.0206	[4] (2017)
	Recovery of full-count PET images from low-count images	35 patients	NA	PET/MR (¹⁸ F-FDG)	PET	MAPE = 4.99 ± 0.68 PSNR = 31.55 ± 1.31 SSIM = 0.9513 ± 0.0154	[5] (2020)
	Improvement of PET resolution	30 patients	Public dataset	PET/CT (¹⁸ F-FDG)	PET	NRM-SR = 2.547–2.642 SSIM = 0.926–0.941 PSNR = 34.17–35.57	[6] (2020)
	Improvement of PET resolution	30 patients	Phantom + Public dataset	NA	PET and MR	CNR = 0.41–0.91 SSIM = 0.88–0.98 PSNR = 37.02–38.79	[7] (2020)
	Mitigation of defective block detectors	20 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET sinograms	MS-SSIM = 0.947–0.984	[8] (2019)
	Improvement of PET resolution	Phantom	Phantom	PET/CT (¹⁸ F-FDG)	Simulated PET low-resolution sinograms	Images with super-resolution sinograms have higher PSNR, MS-SSIM and CRC than those from comparison sinograms with the same crystal bin size, independent of noise	[9] (2018)
	Improvement of PET resolution	Phantom	Phantom	PET/CT (¹⁸ F-FDG)	Flood maps	Average DOI resolution = 2.99 mm (8 × 8 array) and 3.14 mm (4 × 4 array) FWHM	[10] (2020)
	Recovery of partial-ring PET images	Phantom	Phantom	PET/CT (¹⁸ F-FDG)	PET and PET sinograms	MSE = reduction of 91.7% SSIM = 0.993–0.998 RC = 79–94%	[11] (2019)
	PET image reconstruction	NA	Phantom + Retrospective	PET/CT (¹³ C-DASB and ¹⁸ F-FDG)	PET	SSIM = 0.496	[12] (2018)
	PET image reconstruction	Phantom	Phantom	PET/CT (¹⁸ F-FDG)	PET	Better performance than existing methods regarding contrast recovery vs noise trade-offs	[13] (2019)

Table 2 (continued)

Research focus	Aim	Study Population	Study Design	Imaging modality (radiopharmaceutical)	CNN Input	Main results (mean \pm SD or range)	References (publication year)
Technical	AC maps generation	70 patients	Retrospective	PET/MR (^{18}F -FDG)	MR	R2 and RMSE were respectively 1.0 and 253.5 for PET _{U-net} vs. PET _{CTAC} . Bias on SUV measurements: IQ75%–IQ25% = $-0.2 \pm 5.6\%$	[14] (2019)
	AC maps generation	26 patients	NA	PET/MR (^{18}F -FDG and ^{68}Ga -PSMA-11)	MR	RMSE bone lesions = 2.68% RMSE soft tissue lesions = 4.07%	[15] (2017)
	AC maps generation	18 patients	Retrospective	PET/MR (^{18}F -FDG)	MR	Mean Square Root Error = 4.9%	[16] (2018)
	AC maps generation	40 patients	Retrospective	PET/CT (^{18}F -FP-CIT)	MLAA-based activity and attenuation maps	Dice similarity coefficient = 0.79	[17] (2018)
	AC maps generation	100 patients	Retrospective	PET/MR (^{18}F -FDG)	MLAA-based activity and attenuation maps	Dice similarity coefficient = 0.77	[18] (2019)
	AC maps generation	180 patients	Retrospective	PET/CT (^{18}F -FDG, ^{18}F -DOPA, ^{18}F -Flortaucipir, ^{18}F -Flutemetamol)	PET	Outstanding and consistent performance regardless of the radiotracer used resulting in less than 8% absolute SUV bias compared to the standard method (more than 16% bias)	[19] (2020)
	AC maps generation	76 patients	Retrospective	PET/MR (^{11}C -WAY-100635 and ^{11}C -DASB)	MR	Global biases in static PET = $-0.49\% \pm 1.7\%$, and $-1.52\% \pm 0.73\%$ for ^{11}C -WAY-100635 and ^{11}C -DASB, respectively	[20] (2018)
	TOF estimation	Phantom	Phantom	PET/CT (^{18}F -FDG)	Coincident Waveforms	Improvement of time resolution = 20–23%	[21] (2019)
	Dual tracer PET reconstruction	Phantom	NA	PET/CT (^{11}C -FMZ & ^{11}C -DTBZ)	PET sinograms	The bias can be as low as 0.0137 and variance can be as low as 0.0002	[22] (2019)
	Improve the fusion of the complementary information in PET-CT	50 patients	Retrospective	PET/CT (^{18}F -FDG)	PET and CT	Accuracy = 99% Dice score = 0.6385	[23] (2019)

Table 2 (continued)

Research focus	Aim	Study Population	Study Design	Imaging modality (radiopharmaceutical)	CNN Input	Main results (mean ± SD or range)	References (publication year)
Bones Segmentation		146 patients	Retrospective	PET/CT (¹⁸ F-FDG, ¹⁸ F-choline, ¹⁸ F-NaF)	CT	Median SDIs = 0.86 (Th7), 0.85 (L3), 0.88 (sacrum), 0.84 (7th rib) and 0.83 (sternum)	[24] (2019)
Voxel Dose Prediction		Ten patients	Retrospective	PET/CT (⁶⁸ Ga-NOTA-RGD)	PET and CT	Average organ absorbed dose errors = 1.07%	[25] (2019)

AC attenuation correction; CNN convolutional neural network; CRC contrast recovery coefficients; CT computed tomography; ¹¹C-DASB ¹¹C-3-amino-4-benzonitrile; ¹¹C-DTBZ ¹¹C-di-hydro tetraenzazine; ¹¹C-FMZ ¹¹C-flumazenil; DOI depth of interaction; ¹⁸F-FDG 18F-fluorodeoxyglucose; FWHM full width at half maximum; MLAA maximum likelihood reconstruction of activity and attenuation; MR magnetic resonance; MSE mean squared error; MS-SSIM multiscale structural similarity; NA not available data; NMSE normalized mean squared error; NRMSE normalized root mean square error; NRM-SR no-reference quality metric for super-resolution approach; PET positron emission tomography; PSNR peak signal-to-noise ratio; R2 coefficient of determination R2; RC recovery coefficient; RMSE root mean square error; SDIs Sørensen-Dice indexes; SSIM structure similarity; TOF time-of-flight

resolution were investigated by Song et al. [6, 7]. They conducted two investigations to improve PET image resolution.

They developed a self-supervised super-resolution technique (SSSR) for PET, based on dual generative adversarial networks (GANs). Inputs for the SSSR were: a low-resolution PET image, a high-resolution anatomical magnetic resonance image (MR), spatial information (axial and radial coordinates), and a high-dimensional feature set coming from an adjunct CNN. Good performance was achieved in image quality, peak signal-to-noise ratio, structural similarity index, and contrast-to-noise ratio.

Subsequently, the group designed, implemented, and validated several CNN architectures for super-resolution (SR) PET imaging, including shallow and deep varieties. They used the low-resolution PET with its high-resolution anatomical counterpart (e.g. a T1-weighted MR image) as input images. CNN outperformed penalized deconvolution and partial volume correction. The superior performance was demonstrated qualitatively (edge and contrast recovery) and quantitatively (PSNR, SSIM, and on the contrast-to-noise ratio (CNR)).

Whiteley et al. proposed a sinogram repair method based on a CNN able to mitigate the effects of malfunctioning of block detectors, which usually leads to a decreased sensitivity. The proposed method outperformed previously tested methods [8].

Thin-pixelated crystals provide high spatial resolution, but PET systems with such characteristics are not widely available. Hong et al. proposed a data-driven, single-image super-resolution (SISR) approach to enhance the PET image resolution and noise property for PET scanners with large pixelated crystals. They achieved fair image resolution and noise property results (comparable image qualities with four times larger crystals) [9].

Low spatial resolution in pre-clinical and clinical PET scanners with an extended field of view (FOV) can be related to the parallax error, which increases the uncertainty estimation of the annihilation position. Zatcepin et al. developed two DL-based algorithms to estimate depth-of-interaction (DOI) in depolished PET detector arrays, a dense NN and a CNN, and multiple linear regression (MLR) based methods. Tests were performed on an 8 × 8 array of 1.53 × 1.53 × 15 mm³ crystals and a 4 × 4 array of 3.1 × 3.1 × 15 mm³; both coupled to a 4 × 4 array 3 × 3 mm³ silicon photomultipliers. DL-based methods performed better than MLR-based methods and other conventional linear methods, achieving an average DOI resolution of = 2.99 mm (8 × 8 array) and 3.14 mm (4 × 4 array) full width at half maximum (FWHM) [10].

Incomplete projection data lead to artefacts in the reconstructed image. Liu et al. developed a CNN-based method for the recovery of partial-ring PET images. In this study, 20 digital brain phantoms were used in the Monte Carlo

simulation toolkit, SimSET, to simulate full-ring PET scans. The CNN achieved good performance in terms of mean squared error (MSE), structural similarity (SSIM) index and recovery coefficient (RC), showing the potential to recover partial-ring PET images [11] successfully.

As far as PET image reconstruction is concerned, Kim et al. proposed a denoising CNN-based method integrated within the iterative PET reconstruction framework. The algorithm outperformed conventional methods based on total variation (TV) and non-local means (NLM) penalties [12].

Finally, Gong et al. trained a deep residual CNN to improve PET image quality using the existing inter-patient information embedded in the NN. Additionally, the algorithm was integrated into the iterative reconstruction framework. The proposed approach outperformed neural network denoising methods and other conventional methods (the Gaussian filter and penalized reconstruction methods) [13].

Technical applications

One of the most critical technical challenges in PET/MR is an accurate PET attenuation correction (AC) estimation. Seven [14–20] out of the twelve included studies investigated the potential role of CNN-based methods in the field of AC.

Blanc-Durand et al. [14] proposed generating the AC-maps from Zero Echo Time (ZTE) MRI images. Three different methods were compared to the reference CT-based AC map: a single-head atlas-based method, a ZTE-segmentation based method and a CNN-based method with a U-Net architecture. The best performance was achieved by the U-Net AC method that showed the lowest bias, the lowest inter-individual, inter-regional variability, with a negligible impact on brain metabolism estimation.

Leynes et al. [15] proposed a DL model to directly synthesize PseudoCT images from patient-specific multiparametric MRI (Dixon MRI) and a proton-density-weighted ZTE MRI, named ZEDD-CT. The proposed CNN-based method achieved a $4\times$ and $1.5\times$ reduction in root-mean-squared-error (RMSE) quantification of bone and soft tissue lesions, respectively.

Bradshaw et al. [16] evaluated DL's potential use for PET/MR attenuation correction in the pelvis using diagnostic MRI. They found that the DL-based approach outperformed the one using dedicated attenuation correction MRI sequences, shortening the scanning time.

Hwang et al. [17, 18] in 2018 and 2019 investigated different DL-based approaches to improve the simultaneous reconstruction of activity and attenuation in PET imaging based on maximum likelihood reconstruction of activity attenuation (MLAA) approach. In the first one, they proposed three different CNN architectures to learn CT-based

attenuation map from the MLAA-generated activity distribution and attenuation map. The three proposed models were: Convolutional Autoencoder (CAE), U-Net, hybrid CAE and U-net. The hybrid architecture yielded the best results with a Dice similarity coefficient of 0.79 in the bone and 0.72 in the air cavity.

The second study aimed to improve total-body PET/MRI attenuation correction and compare with the Dixon-based four-segment method. The average Dice similarity coefficient (bone regions) between μ -CNN and μ -CT was 0.77, thus providing a reliable attenuation map.

Arabi et al. trained one CNN to generate PET-AC images (PET-DLAC) from PET-non-AC images. They evaluated the quantification accuracy in four datasets (^{18}F -FDG, ^{18}F -DOPA, ^{18}F -Flortaucipir, and ^{18}F -Flutemetamol) PET-CTAC images as reference. DLAC achieved less than 9% absolute SUV bias within each tracer dataset, but it appeared susceptible to outliers [19].

Spuhler et al. developed a CNN-based method to generate patient-specific transmission data from T1-weighted MRI for PET/MRI neuroimaging; they assessed both static and dynamic reconstructions. Good accuracy was shown for both reconstructions by the DL approach. The mean bias was $-1.06 \pm 0.81\%$ for generated transmission data [20].

Berg et al. proposed a CNN-based method to estimate the TOF PET using pairs of digitized detector waveforms for a coincident event as input. A 20% and 23% improvement in time resolution vs leading-edge discrimination and vs constant fraction discrimination, respectively, was achieved [21].

Xu et al. [22] explored the potential of a 3D CNN-based method for dual-tracer PET images reconstruction. They developed a hybrid loss-guided DL-based framework using sinogram data. The proposed algorithm outperformed comparison methods, successfully recovering the distribution of lower total counts. The proposed approach was promising for two tracers' simultaneous imaging, even for tracers labelled with the same isotope.

Kumar et al. proposed a CNN-based method to improve PET-CT fusion. The proposed method encoded modality-specific features and then used them to derive a spatially varying fusion map quantifying the relative importance of each modality's feature across different anatomical regions. Consequently, fusion maps were multiplied by the modality-specific feature maps to obtain a representation of the complementary multi-modality information at different locations. The DL method ability to detect and segment multiple regions was evaluated and compared to reference techniques for multi-modality image fusion (fused inputs, multi-branch, and multi-channel techniques) and segmentation. The developed CNN resulted in a significantly higher foreground detection accuracy and Dice score [23].

As a first step in developing an automated method able to quantify skeletal tumour burden in PET/CT, Belal et al. developed a CNN-based method for bone segmentation and compared its performance with manual segmentations made by an experienced physician. Sørensen-Dice index (SDI) was used to measure the segmentation accuracy. The average volume difference (volume difference/mean volume) between the two segmentations was 5–6% and < 3% for the vertebral column and ribs, and for other bones, respectively [24].

Lee et al. proposed a CNN-based method for voxel dose prediction from PET and CT image patches used as inputs in the radiotherapy planning setting. The voxel dose rate maps predicted by the CNN were compared with a) the ground truth from direct Monte Carlo and b) dose rate maps generated from voxel S-value (VSV) kernel convolution method. Results showed good agreement with the ground truth (voxel dose rate errors = $2.54\% \pm 2.09\%$). Significant improvements were achieved in comparison to the conventional dosimetry approaches [25].

Clinical studies

Summary of the studies' features and main results is provided in Table 3.

Brain and head and neck cancer

In medical imaging, segmentation is a common task; it is used for radiotherapy planning, treatment response assessment and prognostic parameters calculation. An automated approach (full 3D U-Net CNN) for brain lesion segmentation from ^{18}F -FET PET images in patients showing different glial tumours was tested. The authors demonstrated promising performance: a Dice similarity coefficient (DSC) up to 0.8231 was obtained [26].

Radiation therapy is one of the most effective therapeutic strategies in head and neck cancer patients. Treatment success strongly relies on a precise delineation of gross tumour volume (GTV) on medical images. Huang et al. developed and verified an automated GTV segmentation method based on CNN and PET-CT images. Dice similarity coefficient (DSC) of GTV of the proposed method was higher than the previously described automated approaches [27].

Olin et al. described further steps forward radiotherapy planning using CNN-based methods. They tested the feasibility of an automated “one-stop-shop” radiotherapy planning framework using PET/MR data. All dosimetric parameters of the synthetic CT-based dose plans resulted within $\pm 1\%$ of the conventional dose plans [28].

Lymph node staging is crucial since it influences both the overall survival and the probability of distant metastases. Chen et al. combined radiomics, and DL approaches

to classify lymph nodes. They designed a “many-objective radiomics” (MaO-radiomics) model and a 3-dimensional convolutional neural network (3D-CNN). The algorithm fully utilized spatial contextual information and fused the outputs through an evidential reasoning approach. The hybrid method showed an accuracy of 0.88 [29].

Lymphoma

CNN-based methods in lymphoma patients provided good performance in detection and characterization of ^{18}F -FDG-avid lesions. In particular, in the study by Capobianco et al. the CNN-based total metabolic tumour volume (TMTV) was compared to the reference TMTV in terms of prognostic value for progression-free survival (PFS) and overall survival (OS). CNN-derived TMTV was significantly correlated with the reference TMTV ($r = 0.76$; $p < 0.001$). In 280 patients, 6737 ROI_{PARS} (PARS = PET assisted Reporting System) were obtained applying the CNN-based method, while the ROI_{REF} were 7996. The CNN yielded 3317 true negatives, 2399 true positives, 589 false negatives and 432 false positives. Both TMTV resulted in predictive of PFS and OS [30].

Sadik et al. developed a DL-based method to automatically quantify the uptake in the liver and mediastinal blood pool needed to determine the Deauville score, as the first step towards an automated treatment response evaluation. Good accordance between the proposed method and experienced radiologists was achieved [31].

Sites of physiological ^{18}F -FDG uptake and normal excretion (sFEPUs) can interfere in the interpretation of abnormal PET findings and reduce the sensitivity. Bi et al. focussed on the potential use of a CNN-based method—a multiscale superpixel-based encoding (MSE) in sFEPUs identification. Their method outperformed other existing methods in the classification of sFEPUs with average F-score of 0.9173 [32].

Lung cancer

Among the included articles, 7/40 clinical studies investigated the potential of CNN-based methods in lung cancer patients. In lung lesion detection false positives (FPs)-reduction was a central issue [33, 34]. Interestingly, Teramoto et al. developed an FPs reduction method by incorporating CNN into FPs reduction technique that used shape features from PET images' CT and metabolic features. The proposed ensemble technique showed a 90% sensitivity and 4.9 FPs/case [33].

Zhao et al. developed a multi-modality segmentation method relying on FDG uptake and CT information for tumour delineation. They demonstrated that the proposed PET/CT CNN-based method achieved a significant

Table 3 Summary of the characteristics of the 40 selected clinical studies

Research focus	Aim	Study population	Study design	Imaging modality (radiopharmaceutical)	Input modality	Main results	References (publication year)
Head and neck cancer	Glioma detection and segmentation	37	Retrospective	PET/CT (¹⁸ F-FET)	PET	DSC = 0.82	[26] (2018)
	GTV contouring	22 patients	NA	PET/CT (¹⁸ F-FDG)	PET and CT	GTV median DSC = 0.79, sensitivity = 76%, precision = 0.79	[27] (2018)
	Radiotherapy planning	11 patients	Prospective	PET/MR (¹⁸ F-FDG)	MR	Dice coefficient in the comparison of volumes delineated on PET/synthetic CT and PET/CT = 0.87 to 0.99	[28] (2020)
Lymphoma	Lymph node staging	59 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET and CT	Accuracy of hybrid (radiomics and CNN) model = 88%	[29] (2020)
	Diagnosis—TMTV prognostic value	280 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET and CT	85% accuracy, 80% sensitivity, 88% specificity; both TMTV's predictive of PFS and OS	[30] (2020)
	Calculate the liver and blood pool SUV for Deauville score	86 patients	Retrospective	PET/CT (¹⁸ F-FDG)	CT	Mean difference (AI vs radiologists) in liver SUV _{median} was 0.2 and in Mediastinum SUV _{median} was 0.2	[31] (2019)
	Identification of sFEPUs	11 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET	Better accuracy to existing methods (F score = 92%)	[32] (2016)
Lung cancer	Improved FP-reduction method for the detection of pulmonary nodules	84 patients	NA	PET/CT (¹⁸ F-FDG)	PET and CT	The sensitivity of detection was 90%, with 4.9 FPs/case	[33] (2016)
	Diagnosis	70 Patients	NA	PET/CT (¹⁸ F-FDG)	PET	Precision = 0.9 Recall = 1 F score = 0.95	[34] (2019)
	Diagnosis—segmentation	84 patients	NA	PET/CT (¹⁸ F-FDG)	PET and CT	DSC = 0.85 (± 0.08)	[35] (2019)
	Staging (T1-T2 vs T3-T4)	472 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET and CT	AUC = 0.83	[36] (2018)
	Staging (N0 vs. N+)	168 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET and CT	Classical ML approaches AUC = 0.92 CNN approach AUC = 0.91	[37] (2017)

Table 3 (continued)

Research focus	Aim	Study population	Study design	Imaging modality (radiopharmaceutical)	Input modality	Main results	References (publication year)
Esophageal cancer	Staging (N) and metastases development prediction	264 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET	Sensitivity = 74% ± 32%, 45% ± 8% Specificity = 84% ± 16%, 79% ± 6% Accuracy = 0.80 ± 0.17, 0.63 ± 0.05 for nodal and distant metastases, respectively	[38] (2019)
	Survival Prediction	96 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET and CT	AUC: 0.87 and 0.90 (2-year OS and 5-year OS, respectively)	[39] (2019)
	Response prediction	107 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET	Sensitivity = 81% Specificity = 81%	[40] (2015)
Prostate cancer	Outcome prediction	548 PET scans	Retrospective	PET/CT (¹⁸ F-FDG)	PET	AUC = 0.74 in prediction of 1-year survival	[41] (2019)
	Prognosis	145 patients	Retrospective	PET/CT (¹⁸ F-Choline)	CT	Automated total lesion uptake and the relation between the volume of abnormal voxels and total prostate volume were associated with OS (p = 0.02)	[42] (2019)
Multiple cancer types	Prostate cancer volume calculation	45 patients	Retrospective	PET/CT (¹⁸ F-Choline)	PET and CT	$r = 0.77, p < 0.01$ (correlation between automated measurement and specimens)	[43] (2019)
	Staging (N0 vs N+)	549 patients	Retrospective	PET/CT (⁶⁸ Ga-PSMA)	CT	AUC = 0.95	[44] (2020)
	Restaging	233 patients	Retrospective	PET/CT (¹⁸ F-FACBC)	PET	AUC = 0.97 (2D-CNN; slice-based approach)	[45] (2020)
Multiple cancer types	Detection of brain lesions	289 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET	Accuracy (overall model) = 82%	[46] (2019)
	Nodal SUV _{max} prediction	185 patients	Retrospective	PET/CT (¹⁸ F-FDG)	CT	The predicted SUV _{max} was associated with the real SUV _{max} (beta estimate = 0.83, $p < 0.0001$)	[47] (2018)

Table 3 (continued)

Research focus	Aim	Study population	Study design	Imaging modality (radiopharmaceutical)	Input modality	Main results	References (publication year)
	Diagnosis (detect ¹⁸ F-FDG positive foci, predict anatomic location, interpret findings)	629 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET and CT	Anatomic localization accuracy = 96% (body regions), 8% (organs), 81% (subregions) Suspicious vs non-suspicious CNN AUC range = 0.78–0.99	[48] (2020)
	Diagnosis (benign vs malignant vs equivocal scan)	3485 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET	Accuracy to predict benign, malignant and equivocal whole-body images was 99, 99, and 87.5%, respectively	[49] (2020)
Cervical cancer	Diagnosis -segmentation	50 Patients	NA	PET/CT (¹⁸ F-FDG)	PET	Mean DSC = 0.84	[50] (2020)
Soft tissue cancer	Metastases prediction	48 patients	Retrospective (public database)	PET/CT (¹⁸ F-FDG)	PET and CT	AUC = 0.85	[51] (2019)
Neurology	Alzheimer disease Diagnosis	1042 patients	Retrospective (Public dataset)	PET/CT (¹⁸ F-FDG)	PET	AUC = 0.98	[52] (2019)
	Alzheimer disease Diagnosis	397 patients	Retrospective (Public dataset)	PET/MR (¹⁸ F-FDG)	PET and MR	Accuracy = 93%	[53] (2018)
	Alzheimer disease Diagnosis	339 patients	Retrospective (Public dataset)	PET/CT (¹⁸ F-FDG)	PET	AUC = 0.95	[54] (2018)
	Alzheimer's disease Diagnosis	2145 patients	Retrospective (Public dataset)	PET/MR (¹⁸ F-FDG)	PET and MR	Accuracy = 90%	[55] (2019)
	Diagnosis—predict normalized SUVR	1216 (366 Florbetapir 89 Florbetapen)	Retrospective (Public dataset)	PET/CT (¹⁸ F-Florbetapir and ¹⁸ F-Florbetapen)	PET	Mean absolute error of the composite SUVR was 0.050	[56] (2019)
	Prognosis—predict cognitive decline in MCI patients	492 patients	Retrospective (Public dataset)	PET/CT (¹⁸ F-FDG and Florbetapir)	PET	Accuracy = 84% AUC = 0.89	[57] (2018)
	Diagnosis and Prognosis	596 patients	Retrospective (Public dataset)	PET/CT (¹⁸ F-FDG)	PET	AUC was 0.81 for predicting conversion of MCI to Alzheimer dementia for conversion time of 0–3 years	[58] (2019)

Table 3 (continued)

Research focus	Aim	Study population	Study design	Imaging modality (radiopharmaceutical)	Input modality	Main results	References (publication year)
	Prognosis predict dementia related to Parkinson disease and Alzheimer's disease in MCI patients	1364 patients	Retrospective and Public dataset	PET/CT (¹⁸ F-FDG)	PET	The model could differentiate MCI patients who would convert to AD (AUC = 0.82) and Parkinson's disease with dementia (AUC = 0.81)	[59] (2020)
	Diagnosis classification of MSA, PSP and IPD	920 patients	NA	PET/CT (¹⁸ F-FDG)	PET	98% sensitivity, 94% specificity, for the classification of IPD 97% and 99.5%, for MSA 83% and 98%, for PSP AUC = 0.94, accuracy = 86%	[60] (2019)
	Diagnosis of Parkinson's disease	310 patients	NA	PET/CT (¹⁸ F-FDG)	PET	AUC = 0.94, accuracy = 86%	[61] (2019)
Cardiology	Diagnosis of cardiac sarcoidosis	85 patients	Retrospective	PET/CT (¹⁸ F-FDG)	Polar maps	Sensitivity = 84% Specificity = 87%	[62] (2019)
Other	Evaluation of CBF	32 patients	Retrospective	PET/MR (NA)	MR	Structural similarity index = 0.732–0.854	[63] (2019)
	Cerebellum uptake	49 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET	SUV uptake measurements showed no significant error in slope and intercept	[64] (2020)
	Patient gender identification	6462 patients	Retrospective	PET/CT (¹⁸ F-FDG)	PET MIP	Accuracy = 99%	[65] (2019)

AUC area under the curve; *CBF* cerebral blood flow; *CNN* convolutional neural network; *CT* computed tomography; *DSC* dice similarity coefficient; ¹⁸*F-FDG* 18F-fluorodeoxyglucose; ¹⁸*F-FET* 18F-fluoro-ethyl-tyrosine; *FP* false positive; *GTV* gross tumour volume; *IPD* idiopathic Parkinson's disease; *MCI* mild cognitive impairment; *MIP* maximum intensity projection; *MR* magnetic resonance; *MSA* multiple system atrophy; *N* lymph node; *NA* not available data; *OS* overall survival; *PET* positron emission tomography; *PSP* progressive supranuclear palsy; *SDI* Sørensen Dice index; *sFEPV* sites of physiological ¹⁸F-FDG uptake; *SUV_{max}* maximum standardized uptake value; *SUVR* standardized uptake value ratios; *TMTV* total metabolic tumour volume

performance gain in tumour segmentation compared to other traditional and ML-based methods [35].

CNN-based methods were also explored as a tool to assist staging in lung cancer; Kirienko et al. tested a CNN, developed using both PET and CT, to classify T parameter (T1-T2 vs T3-T4). The AUC of the model resulted in 0.83 [36].

For nodal staging, Wang et al. developed a CNN and compared it with four classical ML methods. CNN showed sensitivity, specificity, accuracy, and AUC of 84%, 88%, 86%, and 0.91. Diagnostic performance was not significantly different among the tested algorithms [37].

The CNN-method developed by Tau et al. was aimed at predicting disease spread at nodal and distant sites in non-small cell lung cancer. CNN-based algorithm accuracy was higher for predicting nodal than distant metastases: 80% and 63%, respectively [38].

Finally, Baek et al. showed that CNNs trained to perform tumour segmentation (with no other information than physician contours) identified survival-related image features with remarkable prognostic value. The estimated AUC was 0.88 (95% CI: 0.80–0.96) to predict 2-year OS [39].

Oesophageal cancer

CNN-based methods in oesophageal cancer therapy response and outcome prediction were evaluated by Ypsilantis et al. [40] and Yang et al. [41]. In the former, three-slices (3S)-CNN outperformed other competing predictive parameters (e.g., SUVmax and radiomic indexes); An accuracy of 73% has been achieved in predicting non-responders and responders from pre-treatment ^{18}F -FDG-PET/CT images [40]. In the latter paper, CNN-based methods provided promising results in identifying patients who died within 1 year from the initial diagnosis; results suggested that the prediction model could identify tumours with more aggressive behaviour. Hence, both studies built solid ground to lead further investigations supporting future personalized management of patients affected by oesophageal cancer [41].

Prostate cancer

Prognosis, prostate cancer delineation, nodal staging, and recurrence were all four topics that included studies using CNN-based prostate cancer molecular imaging methods.

Polymeri et al. evaluated a DL algorithm on ^{18}F -choline PET/CT images of 145 patients for automated cancer assessment (versus manual segmentation) and OS prediction. Good accordance between manual measurements and automated PET/CT biomarkers was shown. Automated PET/CT measures were significantly associated with OS ($p=0.02$) [42].

Mortensen et al. focussed on comparing manual vs automated prostate cancer assessment in terms of ^{18}F -choline

PET derived parameters. The correlation between automated and manual measurement was significant. CNN segmentation provided volume and conventional PET measures similar to manually derived ones. Mean differences (95% CI) were 1.40 (– 2.26 to 5.06), 0.37 (– 0.01 to 0.75), –0.08 (– 0.30 to 0.14), and 9.61 (– 3.95 to 23.17) of volume, SUVmax, SUVmean, and total lesion uptake, respectively [43].

Hartenstein et al. trained three different CNNs to determine ^{68}Ga -PSMA PET/CT lymph node status from CT alone. The best CNN outperformed two experienced radiologists with an AUC of 0.95 and 0.81, respectively [44].

Finally, Lee et al. [45] evaluated the performance of deep learning approaches in detecting abnormal ^{18}F -FACBC uptake in patients with biochemical cancer recurrence of prostate cancer. Two different CNN architectures were used: a 2D-CNN (ResNet-50), which uses single slices (slice-based approach) and a 3D-CNN (ResNet-14), which uses a hundred slices per PET image (case-based approach). The slice-based approach performed much better than the case-based approach (AUC = 0.971 and 0.699, respectively).

Multiple cancer types

Nobashi et al. evaluated the performance of CNN-based approaches to dichotomously classify ^{18}F -FDG PET/CT brain scans of cancer patients as abnormal vs normal obtaining convincing results. An overall model that averaged all built models' probabilities showed the best accuracy of 82% [46].

Shaish et al. investigated whether CNN can predict the SUV_{max} of lymph nodes in patients with cancer. The predicted SUV_{max} resulted associated with the reference SUV_{max} ($p < 0.0001$) [47].

Sibille et al. tested multiple CNN configurations' performance on a large cohort of lung cancer and lymphoma patients to localize and classify uptake patterns on total body ^{18}F -FDG PET/CT images into suspicious vs non-suspicious for cancer. In the classification the AUC varied considerably depending on the imaging modality: CT alone, AUC = 0.78 (95% confidence interval [CI]: 0.72, 0.83); ^{18}F -FDG PET alone, AUC = 0.97 (95% CI 0.97, 0.98); ^{18}F -FDG PET/CT, AUC = 0.98 (95% CI 0.97, 0.99); ^{18}F -FDG PET/CT maximum intensity projection (MIP), AUC = 0.98 (95% CI 0.98, 0.99); and ^{18}F -FDG PET/CT MIP atlas, AUC = 0.99 (95% CI 0.98, 1.00) [48].

Kawauchi et al. tested two CNN-based methods (A and B) to classify lesions into benign, malignant and equivocal. A total of 76,785 MIP images were analysed. In the total-body analysis, Algorithm A achieved 91% (benign), 100% (malignant) and 57.5% (equivocal) accuracy; while Algorithm B showed 99.4% (benign), 99.4% (malignant) and 87.5% (equivocal) accuracy. In the region-based analysis, the

accuracy in the prediction of malignant uptake regions was 97.3% (head-and-neck), 96.6% (chest), 92.8% (abdomen) and 99.6% (pelvis) [49].

Cervical cancer

Chen et al. evaluated the performance of spatial information embedded CNN (S-CNN) in the detection of cervical cancer, a known challenging task related to its proximity to the bladder. The S-CNN output has been processed by a thresholding method combined with prior information, reaching a mean DSC of 0.84 [50].

Sarcoma

The high mortality rate related to distant metastases prompts the need for an early prediction of disease spread in sarcoma patients. Peng et al. compared the performance of their deep multi-modality collaborative learning method to the state-of-the-art methods, achieving the overall best performance in predicting distant metastases risk with the following results: the best AUC value of 0.84, the best accuracy of 85%, the best sensitivity of 92%, the best F1 score of 86%, also a second-best precision of 81%, and a competitive third-best specificity of 79% [51].

Neurology

CNN-based methods were investigated in Alzheimer's Disease (AD) and Parkinson's Disease (PD), in particular, 7/10 [52–58] focussed on AD, 1/10 focussed on both [59] and 2/10 focussed on PD only [60, 61].

For AD diagnosis, Ding et al. studied the performance of their CNN based on InceptionV3 architecture. The algorithm achieved an AUC of 0.98 (95% confidence interval: 0.94, 1.00) in predicting the clinical diagnosis of AD, outperforming imager evaluation [52].

Liu et al. used one CNN to classify patients affected by AD [53]. They built a multiple deep 3D-CNN and an upper high-level 2D-CNN able to automatically learn generic multi-level and multimodal features from multiple imaging modalities. High accuracy (93%) was achieved for classification of AD versus controls, while for classification of Mild Cognitive Impairment (MCI) and controls accuracy was lower (83%), demonstrating that the classification of this status is challenging.

The same group tested a classification framework's performance based on a combination of 2D CNN and recurrent neural networks (RNNs). The algorithm showed an AUC of 0.95 for AD vs normal controls (NC) classification and 0.84 for MCI vs NC classification [54].

Huang et al. proposed a CNN that integrated the multi-modality information from the hippocampal area of both

T1-MR and ^{18}F -FDG PET images. The accuracy was 90% and 87% for controls vs AD, and for controls vs MCI, respectively [55].

Kim et al. investigated amyloid quantification methods via a DL model. They aimed at developing a one-step quantification algorithm for amyloid PET, using images acquired from multiple institutions with different radiopharmaceuticals. The mean absolute errors of the composite SUV ratio of test sets for ^{18}F -Florbetapir and ^{18}F -Florbetaben PET were 0.06 and 0.05, respectively [56].

Choi et al. [57] developed a CNN-based method trained on ^{18}F -Fluorodeoxyglucose and ^{18}F -Florbetapir PET images to predict future cognitive decline in MCI patients. Results showed an accuracy of 84% in the prediction for conversion to AD in MCI patients, while accuracy for classification between AD and healthy subjects was 96%. The same group, Choi et al., developed a DL-based evaluation of cognitive dysfunction (cognitive signature) on both Parkinson and AD. The proposed algorithm discriminated between AD and controls on ^{18}F -FDG PET/CT, achieving an AUC = 0.94. When this model was directly transferred to images coming from MCI subjects to identify those who would have most likely progressed to AD, the AUC was 0.82; while testing the method on images coming from Parkinson disease patients to discriminate the ones with dementia, the AUC was 0.81 [59].

Yee et al. proposed a CNN-based method to generate a probability score along the continuum of AD. The method based on ^{18}F -FDG-PET images showed the limited prognostic value in predicting future conversion to Dementia Alzheimer Type [58].

Zhao et al. proposed a 3D deep CNN for an automated early differential diagnosis on ^{18}F -FDG PET/CT images to discriminate Idiopathic Parkinson's Disease (IPD) from multiple system atrophy (MSA) and progressive supranuclear palsy (PSP). Performance achieved by the CNN-based method was as follows: 98% sensitivity, 94% specificity, 95.5% positive predictive value (PPV) and 97% negative predictive value (NPV) for the classification of IPD; 97% sensitivity, 99.5% specificity, 99% PPV, and 99% NPV for MSA diagnosis; 83% sensitivity, 98% specificity, 90% PPV, and 98% NPV, for the PSP, respectively. Also, saliency maps were illustrated. It is worth to mention that, among the saliency features discovered by the deep learning methods, the midbrain was implied as well, which is a widely accepted pathological region for movement disorders that were not considered in the analysis of ^{18}F -FDG PET/CT images yet [60].

Manzanera et al. [61] investigated a 3D-CNN model's potential role in the differentiation of PD patients from controls on ^{18}F -FDG PET/CT images, achieving good performance AUC of 0.94 on the test set.

Cardiology

Hirata et al. [62] developed a CNN-based method to retrospectively differentiate cardiac sarcoidosis (CS) and non-CS in ^{18}F -FDG PET/CT images of 85 patients (CS = 33). An appropriate diagnosis could help prevent deadly cardiac events occurring in this particular type of patients such as complete heart block, ventricular or atrial arrhythmias, congestive heart failure, and sudden cardiac death. Performance of the CNN-based method with the ReliefF algorithm's introduction achieved a sensitivity and specificity of 84% and 87%, respectively, outperforming the standardized uptake value (SUV)-based classification method and the coefficient of variance (CoV)-based classification method.

Other applications

Cerebral blood flow (CBF) is altered in many neurological diseases. Guo et al. [63] developed a CNN-based method trained to integrate single and multi-delay arterial spin labelling (ASL) and structural MR to predict gold-standard ^{15}O -water PET CBF maps. Significant improvement in image quality and quantification accuracy was achieved. Results showed good performance with a structural similarity index of 0.732 for the multi delay and 0.854 for a single delay.

Xiong et al. evaluated the performance of three different 3D deep CNNs (U-Net, V-Net, and modified U-Net) in the automated measurement of ^{18}F -FDG uptake in the cerebellum. U-Net CNN yielded the best performance with a Dice coefficient of 0.911 and showed no significant slope and intercepted error in the SUV uptake measurement than an independent reference standard [64]. This study demonstrated the potential of deep CNNs in automated SUV measurement of reference regions.

To prevent patient misidentification, Kawauchi et al. [65] developed a CNN-based method to predict patients' sex from ^{18}F -FDG PET/CT images, achieving an accuracy of 99%. The pelvic region was the most crucial region to classify the patients correctly. Moreover, the DL method was also able to predict the age and body weight.

Discussion

Nuclear medicine field has experienced rapid development of AI-based applications in the last 2 years.

The vast majority of included articles (45/63–71%) were published in 2019–2020. CNN-based algorithms have been proposed for a wide range of PET imaging purposes, encompassing technical and clinical objectives. Indeed, machine learning algorithms have been demonstrated to be of value for image quality improvement, attenuation correction (in

particular for PET/MR systems), and automatic extraction of a higher amount of information from raw and processed images. Clinical applications comprised oncology (detection, diagnosis, and prognostication in many cancer diseases), neurology and cardiology, in line with the PET/CT indications.

One of the main challenges for CNN-based algorithms development is the scarcity of the datasets. Augmentation strategies are generally put in place to improve model performance and overcome the overfitting phenomenon, a common problem related to machine learning algorithms. Augmenting the data allows adding variability in the dataset to improve the prediction generalization [66]. The selected technical studies included up to 180 subjects, and clinical studies patient population ranged between 11 and 6462 patients, with a median of 209. Deep learning methods require exponentially larger populations. On the one hand, this is necessary to minimize the effects of overfitting. On the other hand, it allows training an algorithm on a cohort representative of the “real-world” population for which the model is developed. It has been demonstrated that model performance significantly improves with dataset enlargement. When a 1000 samples dataset vs the > 100,000 datasets was used for retinopathy classifier development, the weighted error resulted in 13% vs 7%, respectively [67]. Several studies can be considered proof-of-concept or preliminary investigations because of the limited dataset size, restricting their clinical practice applicability. Large and representative study cohorts are challenging to enrol because of ethical limitations, expense, time requirements, or lack of ground truth. Indeed, retrospective study design rather than prospective is the main one among the selected studies.

Alternative strategies have been implemented to overcome this challenge. Multiple studies focussed on neurological diseases (8/10 selected studies, [52–59]) used image datasets from the Alzheimer disease neuroimaging initiative (ADNI) (<http://adni.loni.usc.edu>). ADNI started in 2004 with an initial ANDI-1 project, followed by ADNI-2 and ADNI-3, to detect and track AD using genetic, biochemical, clinical, and imaging biomarkers. The availability of this open database allows multiple investigators worldwide to study and develop alternative strategies to respond to the need for early identification of the disease and risk stratification. The availability of such public datasets for a wide range of conditions could accelerate the research path. Some initiatives such as The Cancer Imaging Archive (TCIA)—a project funded by the Cancer Imaging Programme of the National Cancer Institute—hosts datasets of different medical imaging types and cancer patients' modalities accessible for public download [68].

Virtual clinical trials (VCTs) (or “in silico” imaging trials or virtual imaging trials) may constitute an alternative approach to evaluate medical imaging technologies and to

perform clinical trials. Within VCTs in medical imaging, the investigators may create models of humans, synthetic datasets, simulate imaging scanners, design and use interpretation models, and emerge from our review especially for technical investigations [9, 11, 13]. This technology is at its early stage for clinical applications. These approaches are challenged by computational complexity, simulation realism, and difficulties for validation but soon may represent an alternative or at least a companion strategy for research in medical imaging [69].

Prior experiences on distributed learning approaches such as Clara platform have been launched to promote collaboration among institutions preserving policy and regulatory aspects (<https://developer.nvidia.com/clara>).

Deep learning applied to images attempts to identify features in an image that could be predictive of the outcome of interest (diagnosis or patient survival) without explicit human programming. On the other hand, radiomics is based on calculating many parameters (histogram and texture features), defined by mathematical formulas, that subsequently are analysed using appropriate statistical methods or ML algorithms to assess their potential diagnostic or predictive value. Interestingly, image mining tools may be based on combining these two strategies (radiomics and CNNs), such as in the study by Peng et al. [51]. They applied and compared handcrafted features (with the random forest for classification and prediction), PET-derived 2D and 3D CNN, and an algorithm integrating in-depth features with texture features to predict distant metastases development in patients affected by soft-tissue sarcoma. The multi-modality (PET/CT) collaborative (radiomics and CNN) learning approach demonstrated the best performance. The proposed combination strategy may overcome single approaches.

Knowledge of the basic principles and awareness of deep learning methods' advantages and limitations should become part of radiologists and nuclear medicine physicians' skills. Restructured training programmes are under development [70]. The availability of educational resources by national and international scientific societies and academia for practising professionals is growing; books [1, 71], journal articles [72–74], meetings [75], webinars, and online resources [76] can be accessed. The integration of AI-based tools into the medical workflow is an up-and-coming trend, and all the professionals working in imaging departments should embrace innovation coming from AI, attend training initiatives and be up to date.

We have to acknowledge that we may have missed some papers from technical and engineering resources. However, we aimed to identify the research trend towards the clinical arena. Secondly, we did not perform a quality assessment of the studies since we intended to include preliminary investigations to identify early trends in CNN-based approaches in PET imaging.

In conclusion, CNN applications for PET/CT and PET/MR are exponentially growing for both technical and clinical purposes. ML algorithms demonstrated promising results with performances equal or even higher compared to conventional approaches. Novel research strategies emerged to face the challenges of ML algorithms development. Introduction of AI-based methods into clinical practice requires dedicated educational initiatives for professionals involved in the medical imaging field to enable a critical appraisal of the advantages and limitations of AI-based tools.

Author contributions MK: Literature search and review, manuscript drafting. MS and MK: Content planning and critical data assessment. MB: Manuscript drafting. FG: Manuscript editing and figures' preparation. ES: Manuscript critical revision and editing. AC: Manuscript critical revision and editing.

Funding The present publication was not supported by any funding from public or private sources.

Data availability Not applicable.

Code availability Not applicable.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest with the present work.

Human and animal rights This article does not contain any studies with human or animal subjects performed by any authors since the present paper deals with a literature review.

References

1. Erickson BJ (2019) Deep learning and machine learning in imaging: basic principles. Artificial intelligence in medical imaging. Springer International Publishing, Cham, pp 39–46
2. Benjamens S, Dhunoo P, Meskó B (2020) The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digit Med* 3:118
3. Zhou L, Schaefferkoetter JD, Tham IWK, Huang G, Yan J (2020) Supervised learning with cycleGAN for low-dose FDG PET image denoising. *Med Image Anal* 65:101770
4. Xiang L, Qiao Y, Nie D, An L, Lin W, Wang Q et al (2017) Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. *Neurocomputing* 267:406–416
5. Spuhler K, Serrano-Sosa M, Cattell R, DeLorenzo C, Huang C (2020) Full-count PET recovery from low-count image using a dilated convolutional neural network. *Med Phys* 47:4928–4938
6. Song TA, Chowdhury SR, Yang F, Dutta J (2020) PET image super-resolution using generative adversarial networks. *Neural Netw* 125:83–91
7. Song T-A, Chowdhury SR, Yang F, Dutta J (2020) Super-resolution PET imaging using convolutional neural networks. *IEEE Trans Comput Imaging* 6:518–528

8. Whiteley W, Gregor J (2019) CNN-based PET sinogram repair to mitigate defective block detectors. *Phys Med Biol* 64:235017
9. Hong X, Zan Y, Weng F, Tao W, Peng Q, Huang Q (2018) Enhancing the image quality via transferred deep residual learning of coarse PET sinograms. *IEEE Trans Med Imaging* 37:2322–2332
10. Zatcepin A, Pizzichemi M, Polesel A, Paganoni M, Auffray E, Ziegler SI et al (2020) Improving depth-of-interaction resolution in pixellated PET detectors using neural networks. *Phys Med Biol* 65:175017
11. Liu C-C, Huang H-M (2019) Partial-ring PET image restoration using a deep learning based method. *Phys Med Biol* 64:225014
12. Kim K, Wu D, Gong K, Dutta J, Kim JH, Son YD et al (2018) Penalized PET reconstruction using deep learning prior and local linear fitting. *IEEE Trans Med Imaging* 37:1478–1487
13. Gong K, Guan J, Kim K, Zhang X, Yang J, Seo Y et al (2019) Iterative PET image reconstruction using convolutional neural network representation. *IEEE Trans Med Imaging* 38:675–685
14. Blanc-Durand P, Khalife M, Sgard B, Kaushik S, Soret M, Tiss A et al (2019) Attenuation correction using 3D deep convolutional neural network for brain 18FFDG PET/MR: Comparison with Atlas, ZTE and CT based attenuation correction. *PLoS ONE* 14:1–12
15. Leynes AP, Yang J, Wiesinger F, Kaushik SS, Shanbhag DD, Seo Y et al (2018) Zero-echo-time and dixon deep pseudo-CT (ZeDD CT): Direct generation of pseudo-CT images for Pelvic PET/MRI attenuation correction using deep convolutional neural networks with multiparametric MRI. *J Nucl Med* 59:852–858
16. Bradshaw TJ, Zhao G, Jang H, Liu F, McMillan AB (2018) Feasibility of deep learning-based PET/MR attenuation correction in the pelvis using only diagnostic MR images. *Tomogr (Ann Arbor, Mich)* 4:138–147
17. Hwang D, Kim KY, Kang SK, Seo S, Paeng JC, Lee DS et al (2018) Improving the accuracy of simultaneously reconstructed activity and attenuation maps using deep learning. *J Nucl Med* 59:1624–1629
18. Hwang D, Kang SK, Kim KY, Seo S, Paeng JC, Lee DS et al (2019) 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* 60:1183–1189
19. Arabi H, Bortolin K, Ginovart N, Garibotto V, Zaidi H (2020) Deep learning-guided joint attenuation and scatter correction in multitracer neuroimaging studies. *Hum Brain Mapp*. <https://doi.org/10.1002/hbm.25039>
20. Spuhler KD, Gardus J, Gao Y, DeLorenzo C, Parsey R, Huang C (2019) Synthesis of patient-specific transmission data for PET attenuation correction for PET/MRI neuroimaging using a convolutional neural network. *J Nucl Med* 60:555–560
21. Berg E, Cherry SR (2018) Using convolutional neural networks to estimate time-of-flight from PET detector waveforms. *Phys Med Biol* 63:1–15
22. Xu J, Liu H (2019) Three-dimensional convolutional neural networks for simultaneous dual-tracer PET imaging. *Phys Med Biol* 64:185016
23. Kumar A, Fulham M, Feng D, Kim J (2020) Co-learning feature fusion maps from PET-CT images of lung cancer. *IEEE Trans Med Imaging* 39:204–217
24. Lindgren Belal S, Sadik M, Kaboteh R, Enqvist O, Ulén J, Poulsen MH et al (2019) Deep learning for segmentation of 49 selected bones in CT scans: First step in automated PET/CT-based 3D quantification of skeletal metastases. *Eur J Radiol* 113:89–95
25. Lee MS, Hwang D, Kim JH, Lee JS (2019) Deep-dose: a voxel dose estimation method using deep convolutional neural network for personalized internal dosimetry. *Sci Rep* 9:10308
26. Blanc-Durand P, Van Der Gucht A, Schaefer N, Itti E, Prior JO (2018) Automatic lesion detection and segmentation of 18F-FET PET in gliomas: a full 3D U-Net convolutional neural network study. *PLoS ONE* 13:0195798
27. Huang B, Chen Z, Wu P-M, Ye Y, Feng S-T, Wong C-YO 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
28. Olin AB, Hansen AE, Rasmussen JH, Ladefoged CN, Berthelsen AK, Håkansson K et al (2020) Feasibility of multiparametric positron emission tomography/magnetic resonance imaging as a one-stop shop for radiation therapy planning for patients with head and neck cancer. *Int J Radiat Oncol* 108:1329–1338
29. Chen L, Zhou Z, Sher D, Zhang Q, Shah J, Pham N-L et al (2019) 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* 64:075011
30. Capobianco N, Meignan MA, Cottreau A-S, Vercellino L, Sibille L, Spottiswoode B et al (2020) Deep learning FDG uptake classification enables total metabolic tumor volume estimation in diffuse large B-cell lymphoma. *J Nucl Med* 62:30–36
31. Sadik M, Lind E, Polymeri E, Enqvist O, Ulén J, Trägårdh E (2019) Automated quantification of reference levels in liver and mediastinal blood pool for the Deauville therapy response classification using FDG-PET/CT in Hodgkin and non-Hodgkin lymphomas. *Clin Physiol Funct Imaging* 39:78–84
32. Bi L, Kim J, Kumar A, Wen L, Feng D, Fulham M (2017) Automatic detection and classification of regions of FDG uptake in whole-body PET-CT lymphoma studies. *Comput Med Imaging Graph* 60:3–10
33. Teramoto A, Fujita H, Yamamuro O, Tamaki T (2016) Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. *Med Phys* 43:2821–2827
34. Zhang R, Cheng C, Zhao X, Li X (2019) Multiscale mask R-CNN-based lung tumor detection using PET imaging. *Mol Imaging* 18:153601211986353
35. Zhao X, Li L, Lu W, Tan S (2018) Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. *Phys Med Biol* 64:015011
36. Kirienko M, Sollini M, Silvestri G, Mognetti S, Voulaz E, Antunovic L et al (2018) Convolutional neural networks promising in lung cancer T-parameter assessment on baseline FDG-PET/CT. *Contrast Media Mol Imaging* 2018:1382309
37. Wang H, Zhou Z, Li Y, Chen Z, Lu P, Wang W et al (2017) comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. *EJNMMI Res* 7:11
38. Tau N, Stundzia A, Yasufuku K, Hussey D, Metser U (2020) Convolutional neural networks in predicting nodal and distant metastatic potential of newly diagnosed non-small cell lung cancer on FDG PET images. *Am J Roentgenol* 215:192–197
39. Baek S, He Y, Allen BG, Buatti JM, Smith BJ, Tong L et al (2019) Deep segmentation networks predict survival of non-small cell lung cancer. *Sci Rep* 9:17286
40. Ypsilantis P-P, Siddique M, Sohn H-M, Davies A, Cook G, Goh V et al (2015) Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. *PLoS ONE* 10:e0137036
41. Yang C-K, Yeh JC-Y, Yu W-H, Chien L-I, Lin K-H, Huang W-S et al (2019) Deep convolutional neural network-based positron emission tomography analysis predicts esophageal cancer outcome. *J Clin Med* 8:844
42. Polymeri E, Sadik M, Kaboteh R, Borrelli P, Enqvist O, Ulén J et al (2020) Deep learning-based quantification of PET/CT

- prostate gland uptake: association with overall survival. *Clin Physiol Funct Imaging* 40:106–113
43. Mortensen MA, Borrelli P, Poulsen MH, Gerke O, Enqvist O, Ulén J et al (2019) Artificial intelligence-based versus manual assessment of prostate cancer in the prostate gland: a method comparison study. *Clin Physiol Funct Imaging* 39:399–406
 44. Hartenstein A, Lübke F, Baur ADJ, Rudolph MM, Furth C, Brenner W et al (2020) Prostate cancer nodal staging: using deep learning to predict 68Ga-PSMA-positivity from CT imaging alone. *Sci Rep* 10:3398
 45. Lee JJ, Yang H, Franc BL, Iagaru A, Davidzon GA (2020) Deep learning detection of prostate cancer recurrence with 18F-FACBC (fluciclovine, Axumin®) positron emission tomography. *Eur J Nucl Med Mol Imaging*. <https://doi.org/10.1007/s00259-020-04912-w>
 46. Nobashi T, Zacharias C, Ellis JK, Ferri V, Koran ME, Franc BL et al (2020) Performance comparison of individual and ensemble CNN models for the classification of brain 18F-FDG-PET scans. *J Digit Imaging* 33:447–455
 47. Shaish H, Mutasa S, Makkar J, Chang P, Schwartz L, Ahmed F (2019) Prediction of lymph node maximum standardized uptake value in patients with cancer using a 3D convolutional neural network: a proof-of-concept study. *Am J Roentgenol* 212:238–244
 48. Sibille L, Seifert R, Avramovic N, Vehren T, Spottiswoode B, Zuehlsdorff S et al (2020) 18F-FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks. *Radiology* 294:445–452
 49. Kawauchi K, Furuya S, Hirata K, Katoh C, Manabe O, Kobayashi K et al (2020) A convolutional neural network-based system to classify patients using FDG PET/CT examinations. *BMC Cancer* 20:227
 50. Chen L, Shen C, Zhou Z, Maquilan G, Albuquerque K, Folkert MR et al (2019) Automatic PET cervical tumor segmentation by combining deep learning and anatomic prior. *Phys Med Biol* 64:085019
 51. Peng Y, Bi L, Guo Y, Feng D, Fulham M, Kim J (2019) Deep multi-modality collaborative learning for distant metastases prediction in PET-CT soft-tissue sarcoma studies. In: *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS. IEEE*, pp 3658–3661
 52. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW et al (2019) A deep learning model to predict a diagnosis of Alzheimer disease by using 18 F-FDG PET of the brain. *Radiology* 290:456–464
 53. Liu M, Cheng D, Wang K, Wang Y (2018) Multi-modality cascaded convolutional neural networks for Alzheimer’s disease diagnosis. *Neuroinformatics* 16:295–308
 54. Liu M, Cheng D, Yan W (2018) Classification of Alzheimer’s disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Front Neuroinform* 12:1–12
 55. Huang Y, Xu J, Zhou Y, Tong T, Zhuang X (2019) Diagnosis of Alzheimer’s disease via multi-modality 3D convolutional neural network. *Front Neurosci* 13:1–12
 56. Kim J-Y, Suh HY, Ryoo HG, Oh D, Choi H, Paeng JC et al (2010) Amyloid PET Quantification Via End-to-End Training of a Deep Learning. *Nucl Med Mol Imaging* 2019(53):340–348
 57. Choi H, Jin KH (2018) Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behav Brain Res* 344:103–109
 58. Yee E, Popuri K, Beg MF (2020) Quantifying brain metabolism from FDG-PET images into a probability of Alzheimer’s dementia score. *Hum Brain Mapp* 41:5–16
 59. Choi H, Kim YK, Yoon EJ, Lee JY, Lee DS (2020) Cognitive signature of brain FDG PET based on deep learning: domain transfer from Alzheimer’s disease to Parkinson’s disease. *Eur J Nucl Med Mol Imaging* 47:403–412
 60. Zhao Y, Wu P, Wang J, Li H, Navab N, Yakushev I, et al (2019) A 3D Deep Residual Convolutional Neural Network for Differential Diagnosis of Parkinsonian Syndromes on 18F-FDG PET Images. In: *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS. IEEE*, pp 3531–3534
 61. Manzanera OM, Meles SK, Leenders KL, Renken RJ, Pagani M, Arnaldi D et al (2019) Scaled subprofile modeling and convolutional neural networks for the identification of Parkinson’s disease in 3D nuclear imaging data. *Int J Neural Syst* 29:1950010
 62. Togo R, Hirata K, Manabe O, Ohira H, Tsujino I, Magota K et al (2019) Cardiac sarcoidosis classification with deep convolutional neural network-based features using polar maps. *Comput Biol Med* 104:81–86
 63. Guo J, Gong E, Fan AP, Goubran M, Khalighi MM, Zaharchuk G (2019) Predicting 15 O-Water PET cerebral blood flow maps from multi-contrast MRI using a deep convolutional neural network with evaluation of training cohort bias. *J Cereb Blood Flow Metab* 40:2240–2253
 64. Xiong X, Linhardt TJ, Liu W, Smith BJ, Sun W, Bauer C et al (2020) A 3D deep convolutional neural network approach for the automated measurement of cerebellum tracer uptake in FDG PET-CT scans. *Med Phys* 47:1058–1066
 65. Kawauchi K, Hirata K, Katoh C, Ichikawa S, Manabe O, Kobayashi K et al (2019) A convolutional neural network-based system to prevent patient misidentification in FDG-PET examinations. *Sci Rep* 9:7192
 66. van Dyk DA, Meng X-L (2001) The art of data augmentation. *J Comput Graph Stat* 10:1–50
 67. Kermany DDS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 172(1122–1131):e1129
 68. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P et al (2013) The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging J Digit Imaging* 26:1045–1057
 69. Abadi E, Segars WP, Tsui BMW, Kinahan PE, Bottenus N, Frangi AF et al (2020) Virtual clinical trials in medical imaging: a review. *J Med Imaging* 7:42805
 70. Wasserman P, Freels P, Szames D, Kurra C, Hernandez M (2020) The technophysics year: transformation of diagnostic radiology’s clinical year as a matter of necessity. *Acad Radiol*. <https://doi.org/10.1016/j.acra.2020.04.045>
 71. Joshi AV (2020) *Machine Learning and Artificial Intelligence*. Springer International Publishing, Cham
 72. Chartrand G, Cheng PM, Vorontsov E, Drozdal M, Turcotte S, Pal CJ et al (2017) Deep learning: a primer for radiologists. *Radiographics* 37:2113–2131
 73. Nensa F, Demircioglu A, Rischpler C (2019) Artificial intelligence in nuclear medicine. *J Nucl Med* 60:29S–37S
 74. Currie G, Rohren E (2020) *Intelligent imaging in nuclear medicine: the principles of artificial intelligence, machine learning and deep learning*. *Semin Nucl Med*. <https://doi.org/10.1053/j.semnuclmed.2020.08.002>
 75. (2020) European Association of Nuclear Medicine October 22 – 30, 2020 Virtual. *Eur J Nucl Med Mol Imaging* 47, 1–753. <https://doi.org/10.1007/s00259-020-04988-4>
 76. AI resources and training (2020). <https://www.rsna.org/en/education/ai-resources-and-training>. Accessed 25 Sep 2020

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