

SYSTEMATIC REVIEW

Open Access



Deep learning for synthetic PET imaging: a systematic mapping review of techniques, metrics, and clinical relevance

Maria Vaccaro¹, Enrico Rosa^{2,3}, Elisa Placidi^{2*} , Alessia Guarnera⁴, Aurelio Secinaro⁵, Carlo Gandolfo⁴, Maria Carmen Garganese⁶ and Antonio Napolitano¹

Abstract

Background Synthetic positron emission tomography (PET) imaging, enabled by deep learning, represents a promising approach to minimize radiation exposure while preserving diagnostic accuracy. However, variability in methodologies, performance metrics, and clinical applications needs to be assessed. This systematic mapping review examines the current state of research in synthetic PET generation, analyzing their methodological frameworks and evaluating the clinical relevance.

Materials and methods A systematic search in Scopus, PubMed, and Google Scholar (2019–2024) identified peer-reviewed studies on deep learning-based synthetic PET. Review articles, conference abstracts, and inaccessible full texts were excluded. Data extraction covered study characteristics, imaging modalities, architectures, and evaluation metrics. Due to study heterogeneity, the risk of bias was not formally assessed. Results were synthesized through descriptive and quantitative analysis.

Results Of the initial 116 studies retrieved, 34 were included, 25 of them (73.5%) on brain/neuro using magnetic resonance imaging, computed tomography, or low-dose PET data to generate full-dose or tracer-specific PET. Common architectures included convolutional neural networks, generative adversarial networks, and U-Nets. Peak signal-to-noise ratio (PSNR) ranged 22.69–56.87 dB, structural similarity index measure (SSIM) 0.38–1.00 and mean absolute error (MAE) 1.37–72.00%. Whole-body applications were less frequent (9/34, 26.5%) but showed improvements in oncologic imaging, in particular for tumor detection and image quality. Despite promising advancements, challenges remain, including limited data availability, variability in tracer uptake, and the lack of standardized evaluation metrics. The absence of large/multicenter datasets limits the generalizability of findings.

Conclusions This review highlights promising advancements in synthetic PET imaging using deep learning, with several studies demonstrating the potential for high-quality image generation and substantially reduced radiation exposure. These developments are particularly significant in pediatric populations, where minimizing radiation dose is crucial to ensure patient safety and long-term health. Nonetheless, methodological variability and limited clinical validation continue to pose substantial challenges. Future research should prioritize the development of standardized

Maria Vaccaro and Enrico Rosa contributed equally to this work.

*Correspondence:

Elisa Placidi

elisa.placidi@policlinicogemelli.it

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

evaluation protocols, the use of larger and more diverse datasets—including pediatric cohorts—and comprehensive real-world clinical validation to support the safe and effective translation of synthetic PET techniques into clinical practice.

Relevance statement Deep learning-based synthetic PET imaging enhances diagnostics while reducing radiation, but requires methodological standardization and clinical validation for broader adoption.

Key Points

- Deep learning can create full-dose PET images with less radiation exposure.
- Neurological applications dominate synthetic PET research, maintaining essential diagnostic detail.
- Challenges include limited datasets and variability in tracer uptake, necessitating further advancements.

Keywords Artificial intelligence, Deep learning, Magnetic resonance imaging, Positron emission tomography, Tomography (x-ray computed)

Graphical Abstract

Deep learning for synthetic PET imaging: a systematic mapping review of techniques, metrics, and clinical relevance

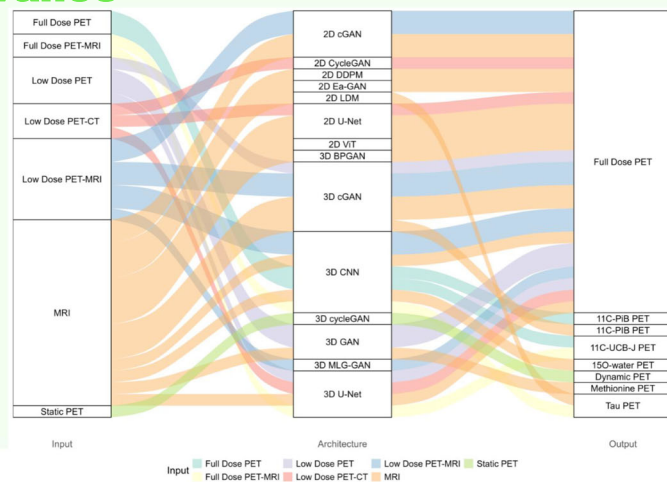
ESRNT EUROPEAN SOCIETY OF RADIOLOGY

- Of 116 studies retrieved, 34 were included, 25 of them (73.5%) on brain/neuro MRI, CT or low-dose PET data to generate full-dose or tracer-specific PET.
- Architectures included CCNs, GANs, and U-Nets.
- Peak signal-to-noise ratio ranged 22.69–56.87 dB, structural similarity index measure 0.38–1.00 and mean absolute error 1.37–72.00%.
- Whole-body applications were only 9/34 (26.5%) but showed improvements in oncologic imaging.
- Challenges remain: limited data availability, variability in tracer uptake, and lack of standardized metrics.

Despite promising advancements, synthetic PET remains an investigational approach



Eur Radiol Exp (2026) Vaccaro M, Rosa E, Placidi E et al; DOI: 10.1186/s41747-025-00651-5



Background

Positron emission tomography (PET) is a key imaging tool in neurology and oncology, providing metabolic and functional information for disease diagnosis and treatment monitoring [1]. However, its widespread clinical use is limited by high costs, long acquisition times, and radiation exposure from radiotracers. These challenges are particularly relevant in vulnerable populations such as children and pregnant women, or patients requiring serial scans [2].

Recent advances in deep learning (DL) have enabled the generation of synthetic PET images from alternative

imaging modalities, such as magnetic resonance imaging (MRI), low-dose PET, or computed tomography (CT), offering a potential solution to reduce radiation exposure while preserving diagnostic accuracy [3]. Various DL architectures, including convolutional neural networks (CNNs), generative adversarial networks (GANs), and U-Net models, have been explored for this purpose [4–7]. These approaches aim to produce synthetic PET images that closely resemble ground-truth PET scans, ensuring high image quality as measured by metrics such as “peak signal-to-noise ratio” (PSNR), “structural similarity index” (SSIM), “mean absolute error” (MAE), and “relative root

mean square error" (rRMSE). Despite promising results, methodological variability, inconsistent evaluation metrics, and limited clinical validation hinder the widespread adoption of synthetic PET imaging.

In the initial stages of medical image synthesis, architectures such as autoencoders and fully convolutional networks were commonly employed, as demonstrated by Nie et al in 2016 [8] and Xiang et al in 2017 [9]. These early models focused on generating synthetic images but faced limitations in their ability to capture complex anatomical and functional relationships. Subsequently, more sophisticated architectures emerged, with U-Net and GANs becoming prominent [10], due to their effectiveness in preserving spatial hierarchies within image data, useful for synthesizing high-quality PET images from input modalities such as MRI, and their capacity to capture the intricate relationships between structural and functional data. For instance, the GLA-GAN model uses global and local modules to enhance synthesis quality, addressing anatomical details at multiple scales [11, 12]. This architectural shift marked a significant improvement in synthetic PET image quality, particularly in the context of full-dose PET synthesis from low-dose PET inputs.

In recent years, state-of-the-art models, such as Vision-Transformers and Diffusion-based architectures, have marked a significant advancement in the next wave of image generative networks. These models represent a turning point from traditional designs as described by Khader et al [13] and Kazerouni et al [14], who highlighted their efficacy in handling complex imaging data, including PET. Vision-Transformer-based methods, in particular, have been demonstrated to incorporate global attention mechanisms, establishing pixel-wise relationships between MRI and PET images [15]. Similarly, diffusion models have shown promising results in PET synthesis, especially in scenarios where high-quality image generation is mandatory [16]. These advanced models exemplify the cutting-edge techniques currently used in medical image synthesis, reflecting the evolution of the field from basic autoencoder-based designs to more sophisticated methods.

Medical image synthesis, particularly synthetic PET, has become a robust tool for avoiding repeated scans with high radiation exposure, particularly for pediatric patients who face increased risks of developing brain cancers and leukemia [17]. The application of synthetic PET is highly relevant in scenarios such as diagnosing degenerative disorders, including Alzheimer's disease, where the cerebral distribution of fluorodeoxyglucose (FDG) in PET scans serves as a crucial differentiating factor [18].

Furthermore, synthetic PET enhances the diagnostic accuracy of cerebrovascular diseases by utilizing MRI-derived cerebral blood flow maps [19]. It also plays a

significant role in developing PET reconstruction algorithms, where synthetic data is used for training and evaluation [20].

Although several reviews have addressed DL approaches for PET image synthesis, to date, no systematic review has comprehensively mapped and analyzed the newest existing literature to assess the performance, clinical feasibility, and diagnostic accuracy of DL-based synthetic PET generation. This review aims to fill this gap by systematically evaluating existing studies, comparing synthetic PET generation from different input images (e.g., MRI-to-PET and CT-to-PET), and identifying the most effective computational approaches. Furthermore, we argue that a formal quality assessment of the papers is a key factor to avoid any scattered results, thus showing a clearer pathway toward optimal solutions. To this end, we provide a structured overview of the field, highlight key limitations, and outline directions for future research and clinical translation while keeping the readers informed on the pitfalls and the quality of the state of the art in PET data generation.

Materials and methods

Literature search

A systematic literature search was conducted using Scopus and PubMed, covering studies published between January 2019 and September 2024. To ensure completeness, Google Scholar was also consulted. The search focused on studies investigating synthetic PET imaging using DL, with keywords such as "synthetic PET" AND "deep learning" applied to titles, abstracts, and keywords. Only English-language peer-reviewed articles were considered. The term 'PET' is used consistently to refer to the synthesized PET image component only, rather than the full hybrid imaging system (e.g., PET/CT or PET/MRI). This reflects the focus of the included studies, which aim to generate synthetic PET images irrespective of the original modality used for acquisition.

The study selection process was conducted independently by two reviewers, who screened studies in two stages: (1) title and abstract screening to exclude irrelevant studies, and (2) full-text review to assess eligibility. Disagreements were resolved through discussion or, if needed, by consulting a third reviewer. Additionally, backward and forward citation searches were performed to identify additional relevant studies, but they did not result in the inclusion of any additional studies.

Eligibility criteria

Studies were included if they: (1) investigated synthetic PET generation using DL models; (2) reported quantitative performance metrics such as PSNR, SSIM, MAE, or

Table 1 Glossary of key concepts, architectures, and evaluation metrics in synthetic PET imaging

Term	Definition/description	Clinical relevance/usage
PET	Positron Emission Tomography. Imaging technique providing metabolic/functional information via radiotracer uptake.	Widely used in oncology and neurology for diagnosis and treatment monitoring.
Synthetic PET	Artificially generated PET images from other imaging modalities using deep learning techniques.	Reduces radiation exposure and scan time; useful in vulnerable populations.
MRI-to-PET/CT-to-PET	Generation of synthetic PET from anatomical images, such as MRI or CT.	Radiation-free or reduced-radiation alternative to PET acquisition.
Low-dose PET	PET imaging acquired with reduced radiotracer dose.	Minimizes radiation; synthetic methods aim to recover full-dose quality.
CNN	Deep learning model effective for spatial feature extraction and noise reduction.	Enhances image quality in low-dose-to-full-dose PET synthesis.
U-Net	Encoder-decoder CNN with skip connections, preserving fine-grained image details.	Common in medical image synthesis for retaining anatomical structures.
GAN	Generative adversarial network—Comprises generator and discriminator networks; generates realistic images.	Used in PET synthesis to improve visual similarity and texture realism.
Vision transformer (ViT)	Attention-based deep model capturing long-range dependencies in image data.	Promising in MRI-to-PET synthesis with global context modeling.
Diffusion model	Probabilistic model iteratively refining noise to generate images.	High fidelity and detail preservation in synthetic PET.
PSNR	Peak signal-to-noise ratio—higher values = lower noise.	Measures image clarity; > 40 dB = high-quality in medical imaging.
SSIM	Structural similarity index—compares luminance, contrast, structure (0 to 1).	Closer to 1 = better structural match with reference image.
MAE	Mean absolute error—average pixel-wise error.	Evaluates intensity accuracy; lower = better.
RMSE/rRMSE	Root Mean Square Error/relative RMSE—global/normalized error measures.	Assesses pixel intensity precision.
Dice score	Measures overlap of predicted and reference regions (0 to 1).	Used in segmentation (<i>e.g.</i> , tumor delineation).
SUV bias	Bias in standardized uptake values—quantifies radiotracer uptake error.	Important for metabolic accuracy in oncology.

MRI Magnetic resonance imaging

rRMSE; (3) used medical imaging modalities (MRI, low-dose PET, or CT) as input; (4) were peer-reviewed and had full-text access. Studies were excluded if they: (1) were review articles, conference abstracts, book chapters, or editorials; (2) did not report quantitative performance evaluations of synthetic PET models; (3) focused on other synthetic imaging modalities unrelated to PET.

Data extraction and analysis

To assess the methodological rigor of the included studies, we employed a checklist inspired by METRICS principles commonly used in AI-centered radiology research [21]. The checklist consists of 11 extracted items covering clinical context, data handling, model development, assessment of DL pipelines, validation, evaluation metrics, reproducibility, and clinical relevance. Each study was evaluated independently across these criteria and assigned a score: 1 for fully satisfied, 0.5 for partially fulfilled, and 0 for unmet. Total scores thus ranged from 0 to 11. Particular attention was paid

to external validation and the availability of open-source code or datasets.

For each included study, key details were extracted and systematically organized to facilitate comparison. These details included: input and output imaging modalities (*e.g.*, MRI to PET, low-dose PET to full-dose PET), DL architectures (*e.g.*, CNNs, GANs, U-Nets, Transformers, Diffusion models), clinical applications (*e.g.*, neurology, oncology, whole-body imaging), evaluation metrics (*e.g.*, PSNR, SSIM, MAE, rRMSE), publication year and references.

A Kruskal-Wallis test was conducted to compare PSNR values between studies that included external validation and those that did not.

Table 1 provides an overview of key terms, architectures, and evaluation metrics relevant to synthetic PET image generation.

Data synthesis and presentation

Both qualitative and quantitative syntheses were conducted to identify trends and assess model performance.

Key quantitative performance metrics were compared across different DL architectures and imaging modalities. A subgroup analysis explored differences in input modalities (MRI, low-dose PET, CT) and their impact on synthetic PET generation, model performance in neurological *versus* whole-body PET applications, and variability in reported performance metrics across studies.

Findings were summarized using descriptive statistics and visually represented through bar plots, alluvial diagrams, and comparative tables.

Results

Overview of the included literature

The search initially retrieved 116 articles. After removing 38 records that were deemed ineligible (21 conference papers, 10 conference reviews, 6 review articles, and 1 book chapter), 78 potentially relevant articles remained. These were further screened based on full-text availability and relevance to the study, resulting in the exclusion of an additional 44 articles. Ultimately, 34 peer-reviewed articles met the inclusion criteria and were selected for analysis.

About the listed studies, 25/34 of them (73.5%) focus on brain-specific applications, and 9/34 of them (26.5%) specifically on whole-body imaging. The key data regarding the input modality, output modality, DL architecture, clinical focus/district and references are synthesized in Table 2.

The study selection process is visually represented in the flowchart of Fig. 1.

A detailed scoring table for all 34 studies is provided in Table 3.

Methodological quality assessment

Among the 34 studies included, all provided a clinical background and described the model architecture. Data transparency was reported in 32 studies (94.1%), while data splitting strategies were reported in 29 studies (85.3%). External validation was performed in 10 studies (29.4%). Regarding evaluation metrics, 32 studies (94.1%) reported them clearly. Robustness assessment of end-to-end DL pipelines was fully evaluated in 10 studies (29.4%). Only 6 studies (17.6%) made their code publicly available. Clinical relevance metrics, such as diagnostic accuracy or expert evaluation, were included in 6 studies (17.6%). Discussion of limitations was reported in 9 studies (26.5%). Total METRICS scores ranged from 4 to 10, with a median score of 6.5. Additionally, of the 34 included studies, 10 (29.4%) performed some form of external validation, while only 6 included clinical evaluation metrics such as diagnostic task performance or reader feedback.

Input modalities and architectures

The selected publications include a wide range of input modalities, such as low-dose ^{18}F -FDG PET (Low-dose PET), MRI, and CT or a combination of these, to generate full-dose ^{18}F -FDG PET (Full-dose PET) or other specific PET modalities. Various DL architectures, including cGANs, CycleGANs, U-Nets, and CNNs, and more recent advancements like vision transformers and denoising diffusion probabilistic models, have been utilized to enhance image synthesis performance, each demonstrating unique capabilities depending on the complexity of the input and the desired output.

Temporal trends and application fields

To visualize the progression of synthetic PET research over time, an analysis was conducted focusing on the use of different DL architectures for brain-specific and whole-body applications shown in Fig. 2. The resulting histogram highlights the distribution of studies from 2019 to 2024, revealing trends in increasing the adoption of such architectures for synthetic PET data creation.

The panel on the left shows the number of articles and the types of architectures employed for synthetic PET in the brain region, while the panel on the right displays the corresponding information for whole-body synthetic PET.

Input–architecture–output relationships

The alluvial plot presented in Fig. 3 illustrates the relationships between different input modalities, architectures, and output types in synthetic PET image generation. The plot highlights how various input data are processed through DL architectures, producing different PET outputs. The flow of connections between these components provides a clear visualization of how different architectures are used for specific types of input and output. This plot serves as a visual summary of the diversity of approaches adopted in synthetic PET research, underscoring the field's ongoing development and innovation in DL methodologies for medical imaging.

As evidence of Fig. 3, MRI stands out as the predominant input modality, underscoring the value of exploring the distribution of MRI input types employed in synthetic PET studies. This distribution is shown in the pie chart presented in Fig. 4.

Quantitative performance metrics

Following the exploration of input modalities, architectures, and output types, we have structured a detailed table (Table 2) that compiles the evaluation metrics for each synthetic PET generation approach, as reported in the reviewed articles. This table provides a breakdown of

Table 2 Evaluation metrics for different synthetic PET generation approaches

Authors [reference number]	PSNR (dB)	SSIM	MAE (%)	NMSE (%)	CNR (%)	rRMSE	PCC	SUV	FID	MAPE (%)	MRAE	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)	MMD
Zhou et al [36]	27.04	0.80	34.27													
Sun et al [54]	32.07	0.99		0.08	0.67											
Xue et al [24]	41.82	1.00				0.12										
Sanaat et al [49]	29.00	0.84				0.40		1.12								
Haggstrom et al [43]	34.69	0.98				0.60										
Wang et al [42]	55.60	0.97				0.16										
Kaplan and Zhou [4]	30.56	0.94				0.27										
Wang et al [47]	25.19	0.98														
Jang et al [33]																
Xie et al [16]	30.92	0.93		0.003												
Zhang et al [52]	27.53	0.93				0.04	0.99									
Shin et al [50]	47.52	0.38				2.38										
Sikka et al [111]	29.33	0.97	1.37													
Wei et al [48]	30.04															
Emami et al [44]	28.45	0.87	46.26						29.17							
Hu et al [12]	27.88	0.89	25.34													
Lin et al [53]	29.42	0.82														
Lee et al [40]		0.90					0.62			0.10						
Wang et al [57]	38.37	0.92	2.14													
Wang et al [38]		0.91				0.16	0.97									
Rajagopal et al [20]		0.94	6.60							0.37						
Chen et al [34]																
Zhang et al [3]																
Kim et al [23]	56.87															
Gao et al [55]	29.10	0.92														
Miao et al [35]		0.70		0.15					10.43							
Abazari et al [37]	30.82	0.82														
Islam and Zhang [51]	32.83	0.77														
Lei et al [25]	30.40		72.00													
Vega et al [22]	22.69	0.93					0.96									
Hussein et al [19]	38.80	0.92				0.04										
Zhang et al [15]	26.92	0.73	3.18													
Takita et al [56]																81.00
Ouyang et al [58]	27.83	0.88				0.15										

This table summarizes the performance of various synthetic PET models using key quantitative metrics. Each row represents a study, linking the employed architecture to its performance on specific metrics. AUC: Area under the receiving operator characteristic curve, CNR: Contrast-to-noise ratio, FID: Fréchet inception distance, MMD: Maximum mean discrepancy, MAE: Mean absolute error, MAPE: Mean absolute percentage error, MRAE: Mean relative absolute error, NMSE: Normalized mean square error, PCC: Pearson correlation coefficient, PSNR: Peak signal-to-noise ratio, rRMSE: Relative root mean square error, SUV: Standardized uptake value, SSIM: Structural similarity index

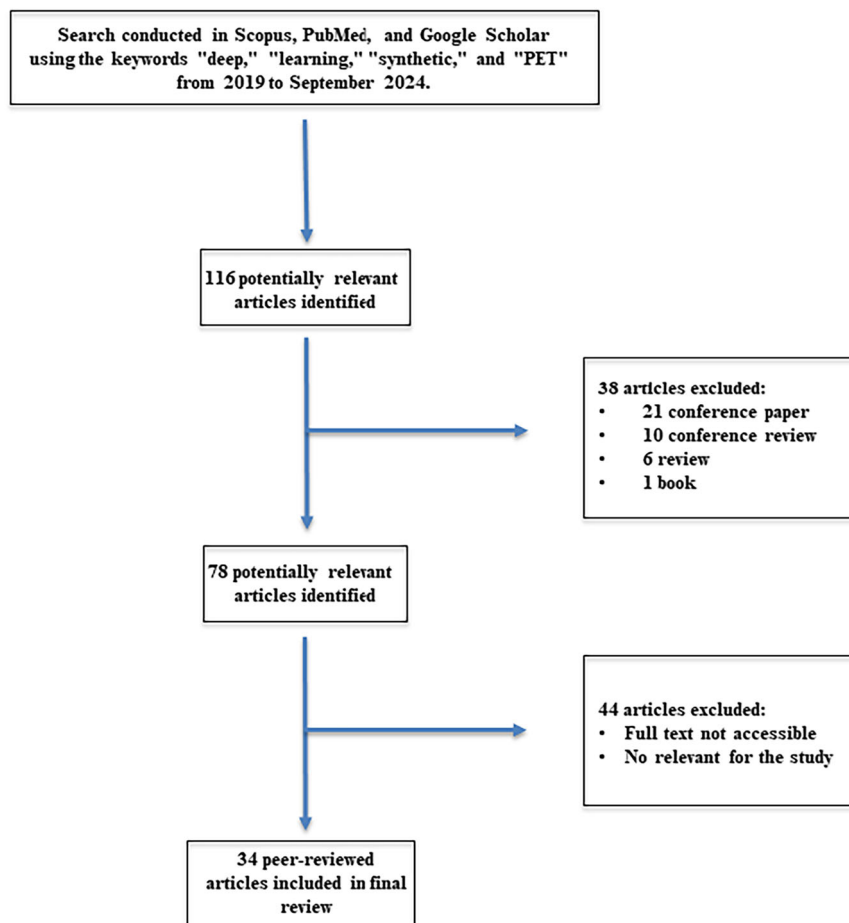


Fig. 1 Flowchart of the literature analysis, including the keywords employed for the search, as well as the inclusion and exclusion criteria applied to select the articles

the performance of different DL models based on key quantitative metrics.

From the table data, the metrics Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Mean Absolute Error (MAE), and relative Root Mean Square Error (rRMSE) were selected for further analysis as they appeared most frequently in multiple articles. PSNR ranges from 22.69 dB (Vega et al [22]) to a maximum value of 56.87 dB in the case of Kim et al [23]. The metric SSIM ranges from 0.38 to 1.00, the maximum found by Xue et al [24]. MAE exhibits a range from 1.37% (Sikka et al [11]) to 72% [25]. The values of rRMSE range between 0.04 (Zhang et al, Hussein et al [15, 19]) and 0.27 [4].

To provide a more comprehensive analysis, we calculated the mean values of the metrics considered in our analysis for each input modality used in synthetic PET image generation. These average values highlight the overall performance of neural networks based on the type

of input data. For instance, when combining PET with another imaging modality, the mean values are: 35.00 dB for PSNR, 0.89 for SSIM, 34.27% for MAE, and 0.28 for rRMSE. Using MRI alone as the input modality yielded a slightly lower PSNR of 31.28 dB, an SSIM of 0.86, an MAE of 22.41%, and an rRMSE of 0.65. In contrast, PET data alone exhibited a higher PSNR of 36.89 dB and an MAE of 34.27%, though the SSIM was 0.77, and the rRMSE value was not determined.

The analysis did not reveal any statistically significant differences ($H = 0.29$, $p = 0.593$), indicating that the presence of external validation was not associated with improved PSNR performance in the available data. The test was performed exclusively on PSNR values, as this was the most consistently reported metric across studies. Other evaluation metrics, such as SSIM, MAE, normalized mean square error, or SUV, were reported too infrequently or in heterogeneous formats, preventing a reliable statistical comparison across groups.

Table 3 Methodological quality assessment of the 34 studies included in the systematic mapping review

Authors [reference number]	Clinical background	Data transparency	Data split	Model architecture	External validation	Evaluation metrics	Comparison with baseline	Robustness assessment of end-to-end deep learning pipelines	Code availability	Clinical relevance	Discussion of limitations	Total score
Haggstrom et al [43]	1	1	0.5	1	0	1	0	0.5	0	0	0.5	5.5
Kaplan and Zhu [4]	1	0.5	0	1	0	1	0	0	0	0	0.5	4
Wang et al [47]	1	1	1	1	1	1	0	1	0	0	1	8
Chen et al [34]	1	1	1	1	0	1	0	0.5	0	0	1	6.5
Wei et al [48]	1	0.5	0.5	1	0	1	0	0.5	0	0	0	4.5
Sanaat et al [49]	1	1	0.5	1	0	1	0	0.5	0	0	0.5	5.5
Shin et al [50]	1	1	0.5	1	0	1	0	0.5	0	0	0.5	5.5
Emami et al [44]	1	1	1	1	0	1	1	0.5	1	0	0	7.5
Islam and Zhang [51]	1	0.5	0	1	0	1	0	0.5	0	0	0.5	4.5
Xue et al [24]	1	1	1	1	0	1	1	0.5	0	0	0	6.5
Wang et al [38]	1	1	1	1	1	1	0	1	0	0	0	7
Zhang et al [52]	1	0.5	0.5	1	0	1	0	0.5	0	0	0	4.5
Sikka et al [11]	1	1	1	1	0	1	0	1	0	0	0	6
Hu et al [12]	1	1	1	1	0	1	0	1	0	0	1	7
Lin et al [53]	1	1	1	1	0	1	0	0.5	0	0	0	5.5
Wang et al [42]	1	1	1	1	1	1	0	1	0	0	1	8
Zhou et al [36]	1	1	1	1	1	1	0	1	0	0	0	7
Sun et al [54]	1	1	1	1	0	1	0	0.5	0	0	0	5.5
Abazari et al [37]	1	1	1	1	0	1	1	0.5	0	0	0	6.5
Zhang et al [15]	1	1	1	1	0	1	0	0.5	1	0	0	6.5

Table 3 continued

Authors [reference number]	Clinical background	Data transparency	Data split	Model architecture	External validation	Evaluation metrics	Comparison with baseline	Robustness assessment of end-to-end deep learning pipelines	Code availability	Clinical relevance	Discussion of limitations	Total score
Jang et al [33]	1	1	1	1	0	1	0	0.5	0	0	1	6.5
Xie et al [16]	1	1	1	1	0	1	0	0.5	0	0	1	6.5
Rajagopal et al [20]	1	1	0.5	1	0	1	0	0.5	0	0	0.5	5.5
Zhang et al [3]	1	1	1	1	1	1	0	0.5	0	0	1	7.5
Kim et al [23]	1	1	1	1	0	1	0	1	0	0	1	7
Gao et al [55]	1	1	1	1	0	1	0	0.5	0	0	0	5.5
Miao et al [35]	1	1	1	1	0	1	0	0.5	0	0	0	5.5
Lei et al [25]	1	1	1	1	0	1	0	1	0	0	0	6
Takita et al [56]	1	1	1	1	1	1	0	1	0	0	1	8
Lee et al [40]	1	1	1	1	1	1	0	0.5	0	0	1	7.5
Wang et al [57]	1	1	1	1	0	1	0	1	1	0	0	7
Vega et al [22]	1	1	1	1	0	1	0	1	1	0	1	78
Hussein et al [19]	1	1	1	1	1	1	0	0.5	0	0	1	7.5
Ouyang et al [58]	1	1	1	1	1	1	1	1	0	1	1	10

The METRICS checklist [59] was used. Each study was evaluated across 11 predefined criteria related to clinical context, data handling, model development, assessment of deep learning pipelines, validation, evaluation, reproducibility, and clinical integration. Scores range from 0 (criterion not satisfied) to 1 (fully satisfied), with 0.5 indicating partial fulfillment. The total score reflects the overall methodological rigor of each study

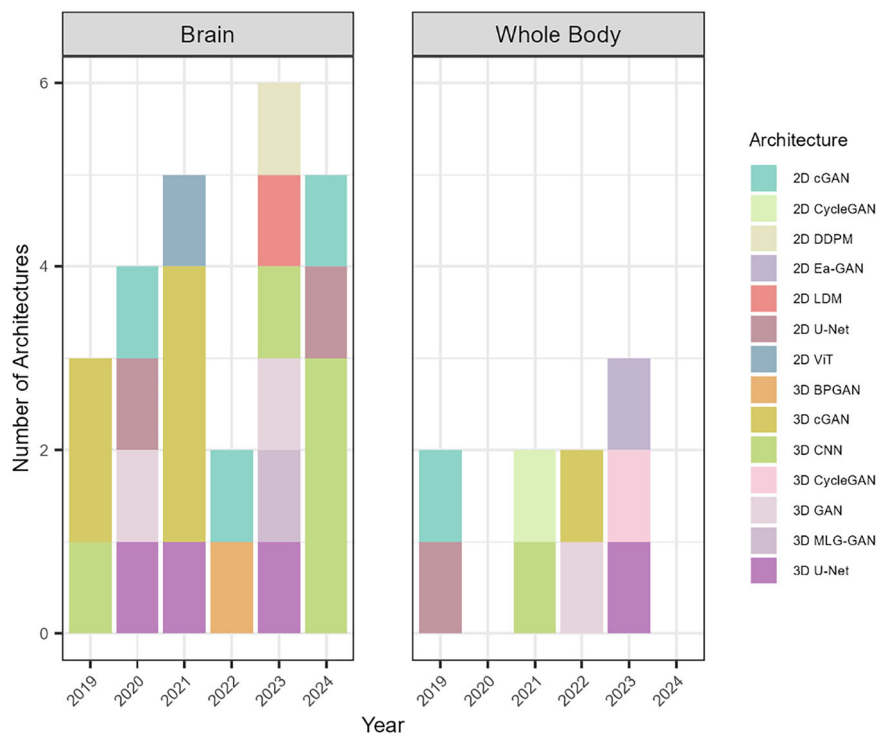


Fig. 2 Chronological overview of deep learning-based networks used for PET image synthesis as reported in the literature. The panel on the left shows the number of articles and the types of architectures employed for synthetic PET in the brain region, while the panel on the right displays the corresponding information for whole-body synthetic PET. 2D, Two-dimensional; 3D, Three-dimensional; BPGAN, Backpropagation generative adversarial network; cGAN, Conditional generative adversarial network; CNN, Convolutional neural network; DDPM, Denoising diffusion probabilistic model; Ea-GAN, Edge-aware generative adversarial network; GAN, Generative adversarial network; LDM, Latent diffusion model; MLG-GAN, Multi-level generative adversarial network; ViT, Vision transformer

Discussion

This review aimed to systematically analyze and compare DL-based synthetic PET imaging techniques, focusing on their network architectures, performance metrics, and clinical applications. Our findings highlight the rapid advancement of synthetic PET imaging, with significant improvements in image synthesis quality and diagnostic potential. However, challenges such as data scarcity, methodological variability, and generalizability remain key obstacles to widespread clinical adoption.

The data illustrate the evolution of synthetic PET methodologies, from early neural network architectures to advanced techniques, that have significantly improved image quality and diagnostic accuracy. These advancements have enabled precise clinical assessments in fields such as oncology and neurology, addressing diverse medical needs with greater efficacy.

As DL continues to advance, further refinement of these methods is expected, driving the broader adoption of synthetic PET imaging in clinical practice and broadening its applicability across a growing spectrum of medical applications.

Evaluating a neural network designed for synthetic PET image generation requires assessing multiple facets, from image fidelity to its clinical applicability. A high-performing network must accurately replicate the fine structural and metabolic details necessary for making a proper diagnosis, while maintaining robust performance across diverse patient populations and scanning protocols. This evaluation goes beyond basic image quality, extending into more nuanced criteria to ensure the neural network is both effective and clinically viable. One of the most important factors in assessing a neural network is its ability to reduce noise while preserving the anatomical features of the image. CNNs, for instance, are particularly effective at enhancing image quality by suppressing noise without blurring key structural elements. This noise reduction is crucial for synthetic PET images generated from low-dose scans, where artefacts can otherwise distort the data and undermine diagnostic utility [26]. Another key aspect is the network's ability to generalize well across various datasets in order to ensure consistent performance, as overfitting to a single dataset can reduce the model's applicability to broader clinical

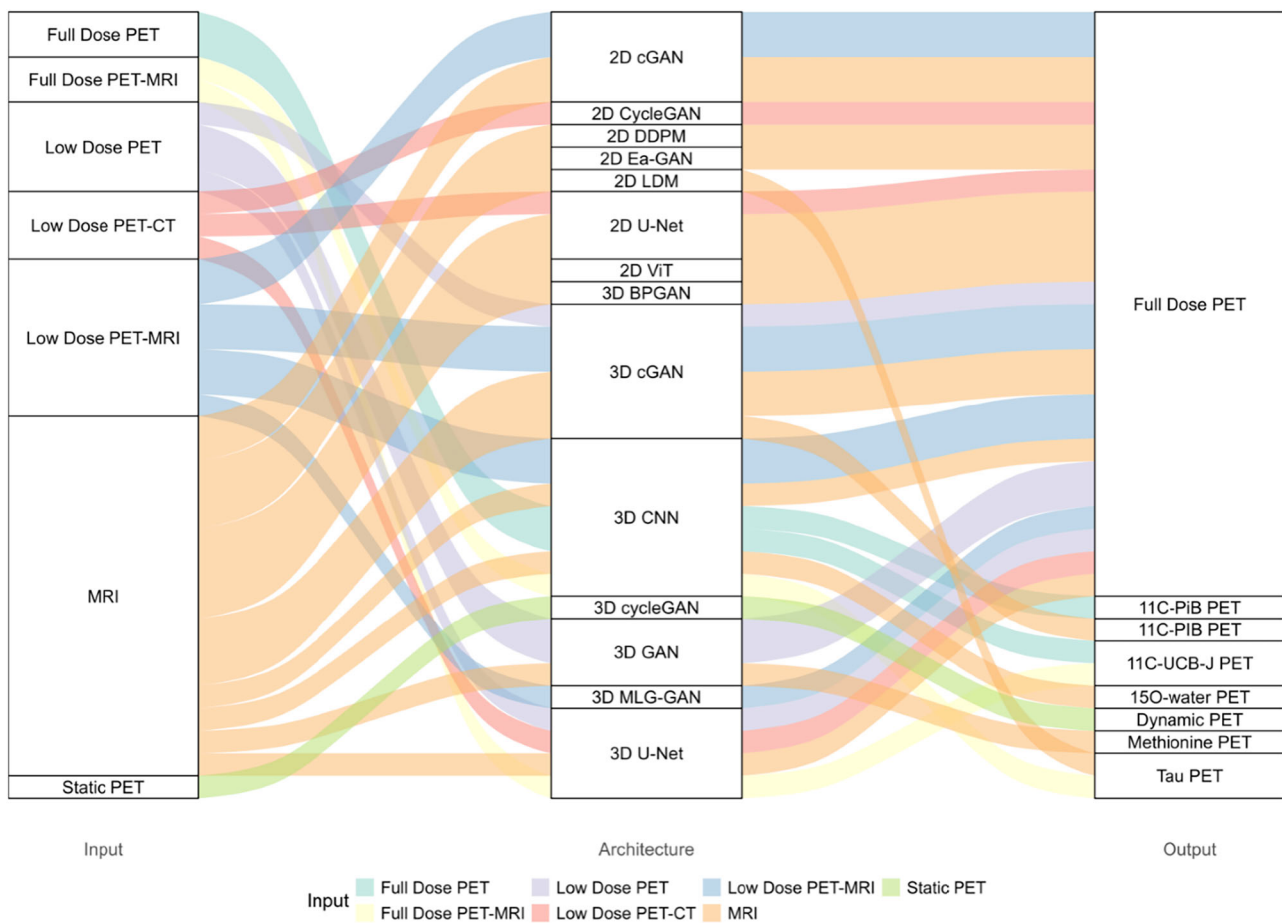


Fig. 3 Alluvial plot depicting the relationships between input modalities, network architectures, and output types for PET data synthesis. 2D, Two-dimensional; 3D, Three-dimensional; BPGAN, Backpropagation generative adversarial network; cGAN, Conditional generative adversarial network; CNN, Convolutional neural network; PET, Positron emission tomography; MLG-GAN, Multi-level generative adversarial network; MRI, Magnetic resonance imaging; CT, Computed tomography; cGAN, Conditional generative adversarial network; DDPM, Denoising diffusion probabilistic model; Ea-GAN, Edge-aware generative adversarial network; GAN, Generative adversarial network; LDM, Latent diffusion model; PET, Positron emission tomography; PiB, Pittsburgh compound B; ViT, Vision transformer

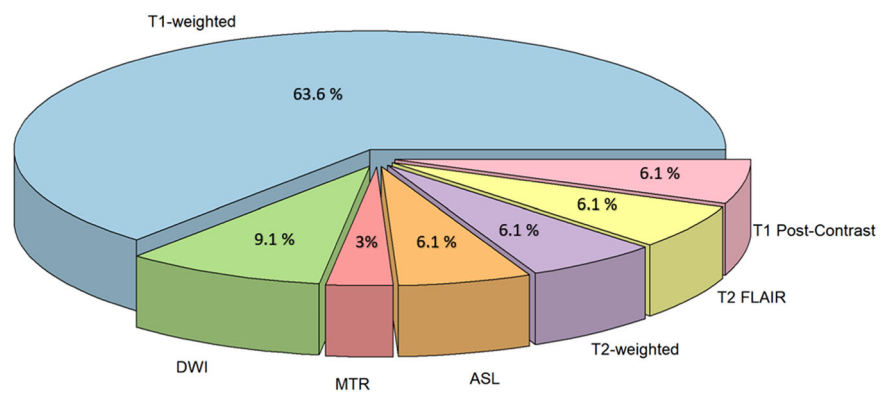


Fig. 4 Distribution of the MRI input modalities/sequences used in synthetic PET studies. ASL, Arterial spin labeling; DWI, Diffusion-weighted imaging; FLAIR, Fluid-attenuated inversion recovery; MTR, Magnetization transfer ratio

environments. Testing across different scanners, imaging protocols, and patient populations is therefore critical to ensure the robustness of the synthetic PET images.

To evaluate neural networks used for synthetic PET image generation, several quantitative metrics are typically employed to assess both image quality and clinical utility. The PSNR quantifies image clarity by comparing signal strength to noise levels. Higher PSNR values indicate superior image quality, making it a reliable measure for comparing synthetic images to their ground-truth counterparts. In research, PSNR values exceeding 40 are often considered indicative of high-quality images in medical imaging applications [27].

Another important metric is the SSIM, which assesses how closely the synthetic image matches the real image in terms of luminance, contrast, and structure. SSIM values range from 0 to 1, with values closer to 1 indicating better image quality and preservation of structural integrity [28].

The MAE quantifies the average pixel-level difference between the synthetic and real PET images, providing insight into the overall accuracy of the model. Lower MAE values indicate better performance, as they reflect a closer match between the synthetic image and the ground-truth data. This metric is particularly useful for assessing how well the model captures small details, which is crucial in medical imaging, where subtle differences can impact clinical decisions [29].

For instance, in synthetic PET image evaluations, MAE is often used alongside other metrics like PSNR and SSIM to ensure that the overall error is minimized and the visual and structural integrity of the image is maintained.

The RMSE is the metric used to evaluate the overall error between predicted and actual images in synthetic PET image generation. It is particularly useful for quantifying the differences in pixel intensity between synthetic images and ground-truth images. A lower RMSE value reflects a higher degree of accuracy in image synthesis, making it an essential measure for ensuring that synthetic PET images closely replicate the original PET scans. Similar to MAE, RMSE is also used with other metrics like SSIM and PSNR in evaluating cross-modality image synthesis, where the task is to maintain precision in intensity and image quality [30].

For segmentation tasks in synthetic PET imaging, the Dice score is commonly used to evaluate how well the predicted regions of interest, such as tumors or lesions, overlap with the ground-truth anatomical structures. This metric is particularly useful in assessing the network's ability to detect and delineate abnormalities accurately. The Dice score measures the similarity between the predicted segmentation and the true regions, with values

closer to 1 indicating a high degree of overlap and accurate segmentation [31, 32].

Lastly, the Standardized Uptake Value Bias (SUV Bias) is useful in evaluating radiotracer uptake in synthetic PET images, especially in oncology. This metric ensures that synthetic images accurately reflect metabolic activity [27].

Incorporating these metrics ensures that the neural networks produce visually accurate synthetic PET images and maintain the clinical relevance required for real-world medical applications.

The results of the METRICS evaluation underscore several key gaps in the current body of work on synthetic PET generation using DL. While most studies achieved good scores in technical aspects such as model design and evaluation metrics, they often lacked reproducibility components like open-source code and external validation. This limits the generalizability and clinical credibility of the proposed approaches. By systematically assessing these aspects, our review not only synthesizes recent advances but also exposes areas that require improvement for future translational research. The availability of a standardized evaluation like METRICS can help raise methodological standards and guide the design of more robust and clinically meaningful AI algorithms.

The PET images generated via advanced DL architectures represent a major leap forward in medical imaging, providing a variety of methods to achieve high-quality synthetic images. As shown in Table 4, various studies have investigated DL architectures for synthetic PET generation, leveraging different input modalities and tracer types across brain and whole-body applications. Figure 2 illustrates the distinct trends between these two domains.

The majority of studies focus on brain imaging, where low-dose PET, MRI, and CT are commonly used as input modalities to generate full-dose PET or specialized PET tracers, such as tau PET [33], 11C-UCB-J PET [34], or dynamic PET scans [35]. This emphasis reflects the growing demand for advanced neuroimaging techniques to enhance diagnosis and disease monitoring in neurological disorders. Additionally, the widespread adoption of synthetic PET in brain imaging highlights its potential to improve diagnostic accuracy while minimizing radiation exposure, a particularly relevant factor in neurology, where high-resolution imaging is essential for understanding disease progression. Although whole-body synthetic PET applications are less prevalent, they remain a key area of investigation, particularly in oncology [36, 37]. The ability to synthesize high-quality PET images for oncological assessments suggests potential clinical benefits, such as enhanced tumor detection and reduced radiation exposure. As DL methods continue to evolve,

Table 4 Summary of studies on synthetic PET image generation (2019–2024)

Input modality	Output modality	Deep learning architecture	Clinical focus/district	Authors [reference number]
Low-dose PET - CT	Full-dose PET	2D U-Net	Whole body	Haggstrom et al [43]
Low-dose PET - MRI	Full-dose PET	2D cGAN	Whole body	Kaplan and Zhu [4]
Low-dose PET - MRI	Full-dose PET	3D cGAN	Brain	Wang et al [47]
Low-dose PET - MRI	Full-dose PET	3D CNN	Brain	Chen et al [34]
MRI	11C-PIB PET	3D cGAN	Brain	Wei et al [48]
Low-dose PET - CT	Full-dose PET	3D U-Net	Brain	Sanaat et al [49]
MRI	Full-dose PET	2D cGAN	Brain	Shin et al [50]
MRI	Full-dose PET	2D U-Net	Brain	Emami et al [44]
Low-dose PET	Full-dose PET	GAN	Brain	Islam and Zhang [51]
Low-dose PET - CT	Full-dose PET	2D CycleGAN	Whole body	Xue et al [24]
Low-dose PET - MRI	Full-dose PET	3D CNN	Whole body	Wang et al [38]
MRI	Full-dose PET	2D ViT	Brain	Zhang et al [52]
MRI	Full-dose PET	3D cGAN	Brain	Sikka et al [11]
MRI	Full-dose PET	3D cGAN	Brain	Hu et al [12]
Low-dose PET - MRI	Full-dose PET	3D cGAN	Brain	Lin et al [53]
Full-dose PET - MRI	11C-UCB-J PET	3D U-Net	Brain	Wang et al [42]
Low-dose PET	Full-dose PET	3D cGAN	Whole body	Zhou et al [36]
Low-dose PET - MRI	Full-dose PET	2D cGAN	Brain	Sun et al [54]
Low-dose PET	Full-dose PET	3D GAN	Whole body	Abazari et al [37]
MRI	Full-dose PET	3D BPGAN	Brain	Zhang et al [15]
MRI	Tau PET	2D LDM	Brain	Jang et al [33]
MRI	Full-dose PET	2D DDPM	Brain	Xie et al [16]
MRI	Full-dose PET	3D U-Net	Whole body	Rajagopal et al [20]
Full-dose PET	11C-UCB-J, 11C-PiB PET	3D CNN	Brain	Zhang et al [3]
Low-dose PET	Full-dose PET	3D U-Net	Brain	Kim et al [23]
Low-dose PET - MRI	Full-dose PET	3D MLG-GAN	Brain	Gao et al [55]
Static PET	Dynamic PET	3D cycleGAN	Whole body	Miao et al [35]
MRI	Full-dose PET	2D EA-GAN	Whole body	Lei et al [25]
MRI	Methionine PET	3D GAN	Brain	Takita et al [56]
Full-dose PET - MRI	Tau PET	3D CNN	Brain	Lee et al [40]
MRI	Full-dose PET	2D U-Net	Brain	Wang et al [57]
MRI	Full-dose PET	2D cGAN	Brain	Vega et al [22]
MRI	15O-water PET CBF	3D CNN	Brain	Hussein et al [19]
MRI	Full-dose PET	3D CNN	Brain	Ouyang et al [58]

The table highlights various input-output combinations used in synthetic PET generation, covering a range of deep learning architectures. The studies are categorized based on the clinical focus/district involved and the imaging modalities employed

CT Computed tomography, cGAN Conditional generative adversarial network, CNN Convolutional neural network, DDPM Denoising diffusion probabilistic model, EA-GAN Edge-aware generative adversarial network, GAN Generative adversarial network, LDM Latent diffusion model, MRI Magnetic resonance imaging, MLG-GAN Multi-level generative adversarial network, PiB Pittsburgh compound B, PET Positron emission tomography, ViT Vision transformer

future research is expected to expand whole-body synthetic PET applications, improving diagnostic workflows and broadening clinical adoption.

The architectures used indicate that 3D cGANs are frequently applied for whole-body imaging, translating low-dose PET images into full-dose PET images. These models capture detailed spatial relationships in PET data while retaining clinically relevant information, as demonstrated in studies by Zhou et al (2022) and Wang et al [36, 38]. The integration of MRI data as input, in

addition to PET images, offers enhanced anatomical detail for improved visualization in synthetic PET [15]. The versatility of 3D cGANs has been shown in malignancy characterization using whole-body MR images [5].

In contrast, 2D cGANs and U-Net architectures are more commonly utilized for brain imaging, where input images are often MRI or CT scans. These architectures demonstrate adaptability to the specific anatomical requirements of brain PET synthesis, emphasizing their suitability for capturing finer details. The capacity of GANs to generate

diverse, realistic samples makes them particularly valuable in medical fields where large public datasets are scarce. For instance, the “globally and locally aware” GAN model incorporates global and local modules to enhance synthesis quality by addressing anatomical details on multiple scales, as reported by Sikka et al in 2021 [11].

Only a limited number of studies explored multi-tracer PET synthesis (e.g., tau, amyloid, synaptic density tracers), despite their growing importance in neurodegenerative disease imaging. Likewise, advanced architectures such as vision transformers and diffusion models, although promising, remain underutilized, with limited evaluation of robustness and generalizability. Most included studies were centered on neuroimaging, reflecting an imbalance in clinical application domains. These observations suggest opportunities for future work in underrepresented areas such as whole-body imaging and tracer-specific synthesis beyond FDG.

The synthesis methods also vary in the tracers used. Many studies focus on synthesizing images for ^{18}F -FDG PET, essential for diagnosing metabolic conditions, especially in neurological disorders [11]. Other studies have explored tracers such as ^{11}C -UCB-J for assessing synaptic density and astrocytosis or gliosis [38] and ^{11}C -PIB for amyloid deposit detection [20]. Another study focuses on ^{15}O - H_2O PET, used to measure cerebral blood flow, which is fundamental in brain function analysis [39]. Tau tracers, as employed in studies by Lee et al and Jang et al [33, 40], play a role in distinguishing patients with rapidly progressive dementia due to Alzheimer’s disease from those with lower neurofibrillary tangle burdens [41]. This growing interest in using synthetic PET to study specific pathologies like Alzheimer’s disease and other neurodegenerative disorders suggests the importance of these biomarkers for early diagnosis and disease monitoring.

The alluvial plot in Fig. 3 reveals that MRI is the predominant input modality for generating various types of synthetic PET images. For this reason, the pie chart in Fig. 4 illustrates the distribution of MRI modalities used in synthetic PET studies, with T1-weighted images dominating at 63.6%. This preference highlights the extensive utility of T1-weighted imaging in PET synthesis tasks due to its superior structural resolution and compatibility with DL models. For instance, studies like Vega et al (2024) demonstrate that synthetic full-dose PET images generated from T1-weighted brain MRI can improve the performance of deep anomaly detection models in neuroimaging applications [22].

Diffusion-weighted imaging (DWI) accounts for 9.1% of inputs, reflecting its importance in capturing micro-structural details that enhance PET synthesis for specific applications, such as neurological diagnostics.

Modalities like Magnetization Transfer Ratio (MTR) represent 3%, reflecting their niche role in addressing specialized imaging needs, particularly in the context of tissue characterization. Meanwhile, Arterial Spin Labeling (ASL), T2-weighted, T2-FLAIR, and T1 post-contrast imaging each comprise 6.1% of the inputs, illustrating their focused application in scenarios requiring enhanced contrast or functional information.

All metrics play a key role in evaluating network performance, but our analysis focuses on the four most common metrics—PSNR, SSIM, MAE, and rRMSE—which are widely recognized in image synthesis literature.

A 3D U-Net architecture [42] achieved the highest performance in PSNR, demonstrating the effectiveness of U-Net-based models for preserving spatial features in synthetic PET images. Similarly, Haggstrom et al [43] used a 2D U-Net and showed noteworthy performance, underscoring the model’s robustness in low-dose to full-dose PET synthesis.

CycleGAN architectures, like the 2D CycleGAN used by Xue et al [24], excel in SSIM, showing a strong ability to retain high-level and fine-grained details.

A 3D cGAN architecture [38], when combined with MRI input for full-dose PET synthesis, performed best in minimizing error, as reflected by its low MAE values. In terms of rRMSE, a 3D CNN architecture [15] outperformed others in brain-region PET image synthesis, demonstrating its ability to maintain high numerical accuracy, essential in diagnosing neurological conditions that require precise spatial representation.

U-Net architectures consistently perform well across multiple studies, confirming their efficiency in preserving spatial hierarchies within image data. Emami et al [44] introduced a frequency-aware U-Net (FREA-UNet), which optimizes synthesis by managing both low- and high-frequency components, resulting in improved image quality and detail preservation. New Transformer-based models, as explored by Li et al (2023) and Shamshad et al (2022), show potential in enhancing feature extraction through spatial attention mechanisms [45, 46].

A comparative analysis of the reviewed studies suggests that U-Net-based architectures, both 2D and 3D, consistently deliver strong results in terms of PSNR and SSIM, particularly for low-dose PET to full-dose PET synthesis in both brain and whole-body applications. Their encoder-decoder structure with skip connections supports the preservation of spatial resolution and anatomical detail. GAN-based models (especially cGANs and CycleGANs) tend to excel in preserving texture and producing visually realistic outputs, often achieving superior SSIM values in MRI-to-PET translation, although they may exhibit instability during training and lack precise control over intensity values.

Transformer-based models and Diffusion models, although less common in the current literature, demonstrate promising results in handling complex multimodal inputs and capturing global contextual features. However, they require substantial computational resources and are still under-evaluated in terms of robustness and clinical applicability. These observations underscore that model selection should be guided not only by architecture but also by the clinical context and data availability.

The clinical deployment of AI-based synthetic PET models raises important regulatory, ethical, and interpretability concerns. As these systems may influence diagnostic decisions, questions of responsibility, transparency, and patient safety become critical. Most reviewed studies do not address model explainability or mechanisms for interpreting output decisions, which limits their acceptability in clinical workflows. Furthermore, regulatory frameworks such as the FDA's guidance on AI/ML-based software as a medical device (SaMD) and the upcoming EU AI Act require rigorous validation, performance monitoring, and risk management before clinical adoption. Meeting these requirements will necessitate greater attention to documentation, human oversight, and post-deployment auditing. Future work should incorporate not only technical performance but also compliance pathways and strategies to ensure AI systems are ethically and legally deployable in real-world settings.

However, further evaluation is required to compare their performance in terms of PSNR and SSIM against well-established architectures like U-Net and GAN-based models. Additionally, diffusion models are emerging as novel approaches for PET image generation, though results are still limited in the current literature.

This review highlights promising advancements in synthetic PET imaging using DL, with several studies demonstrating the potential for high-quality image generation and substantially reduced radiation exposure. These developments are particularly significant in pediatric populations, where minimizing radiation dose is crucial to ensure patient safety and long-term health. Nonetheless, methodological variability and limited clinical validation continue to pose substantial challenges. Future research should prioritize the development of standardized evaluation protocols, the use of larger and more diverse datasets—including pediatric cohorts—and comprehensive real-world clinical validation to support the safe and effective translation of synthetic PET techniques into clinical practice.

Abbreviations

cGAN	Conditional generative adversarial network
CNN	Convolutional neural network
CT	Computed tomography

FDG	Fluorodeoxyglucose
GAN	Generative adversarial network
MAE	Mean absolute error
MRI	Magnetic resonance imaging
PET	Positron emission tomography
PSNR	Peak signal-to-noise ratio
rRMSE	Relative root mean square error
SSIM	Structural similarity index
SUV	Standardized uptake value

Acknowledgements

This work was supported in part by the Italian Ministry of Health with "Current Research funds". No large language models (LLMs), generative AI software, or related tools were used in the preparation, writing, analysis, or editing of this manuscript. All content is the original work of the authors.

Author contributions

MV wrote the initial draft and analyzed the literature data. ER reviewed the literature and wrote the initial draft. EP and AG analyzed the data and revised the initial draft. AS, CG, and MCG provided critical internal reviews and edited the draft to its final version. AN coordinated and led the research group. All authors read and approved the final manuscript.

Funding

This work was supported by "Bando 5x1000 2024" of Bambino Gesù Children's Research Hospital.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details

¹Medical Physics Unit, Bambino Gesù Children's Hospital IRCCS, Rome, Italy. ²UOC Fisica per le Scienze della Vita, Dipartimento di Diagnostica per Immagini e Radioterapia Oncologica, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy. ³eCampus University, Novedrate, Italy. ⁴Functional and Interventional Neuroradiology Unit, Bambino Gesù Children's Hospital IRCCS, Rome, Italy. ⁵Advanced Cardiothoracic Imaging Unit, Bambino Gesù Children's Hospital IRCCS, Rome, Italy. ⁶Nuclear Medicine Unit/Imaging Department, Bambino Gesù Children's Hospital IRCCS, Rome, Italy.

Received: 15 March 2025 Accepted: 17 October 2025

Published online: 09 February 2026

References

1. Artesani A, Bruno A, Gelardi F, Chiti A (2024) Empowering PET: harnessing deep learning for improved clinical insight. *Eur Radiol Exp* 8:17. <https://doi.org/10.1186/s41747-023-00413-1>
2. Shiyam Sundar LK, Gutschmayer S, Maenle M, Beyer T (2024) Extracting value from total-body PET/CT image data—the emerging role of artificial intelligence. *Cancer Imaging* 24:51. <https://doi.org/10.1186/s40644-024-00684-w>
3. Zhang H, Qie Y (2023) Applying deep learning to medical imaging: a review. *Appl Sci* 13:10521. <https://doi.org/10.3390/app131810521>
4. Kaplan S, Zhu Y-M (2019) Full-dose PET image estimation from low-dose PET image using deep learning: a pilot study. *J Digit Imaging* 32:773–778. <https://doi.org/10.1007/s10278-018-0150-3>

5. Pasquini L, Napolitano A, Pignatelli M et al (2022) Synthetic post-contrast imaging through artificial intelligence: clinical applications of virtual and augmented contrast media. *Pharmaceutics* 14:2378. <https://doi.org/10.3390/pharmaceutics14112378>
6. Seyyedi N, Ghafari A, Seyyedi N, Sheikhzadeh P (2024) Deep learning-based techniques for estimating high-quality full-dose positron emission tomography images from low-dose scans: a systematic review. *BMC Med Imaging* 24:238. <https://doi.org/10.1186/s12880-024-01417-y>
7. Zhao K, Zhou L, Gao S et al (2020) Study of low-dose PET image recovery using supervised learning with CycleGAN. *PLoS One* 15:e0238455. <https://doi.org/10.1371/journal.pone.0238455>
8. Nie D, Wang L, Gao Y, Shen D (2016) Fully convolutional networks for multi-modality isointense infant brain image segmentation. In: *Proceedings of the 2016 IEEE 13th international symposium on biomedical imaging (ISBI)*. IEEE, Prague, pp 1342–1345
9. Xiang H, Yuan Y, Zhao Y, Fu Z (2017) License plate detection based on fully convolutional networks. *J Electron Imaging* 26:1. <https://doi.org/10.1117/1.JEI.26.5.053027>
10. Nie D, Trullo R, Petitjean C et al (2016) Medical image synthesis with context-aware generative adversarial networks. <https://doi.org/10.48550/arXiv.1612.05362>
11. Sikka A, Skand, Virk JS, Bathula DR (2021) MRI to PET cross-modality translation using globally and locally aware GAN (GLA-GAN) for multi-modal diagnosis of Alzheimer's disease. <https://doi.org/10.48550/arXiv.2108.02160>
12. Hu S, Lei B, Wang S et al (2022) Bidirectional mapping generative adversarial networks for brain MR to PET synthesis. *IEEE Trans Med Imaging* 41:145–157. <https://doi.org/10.1109/TMI.2021.3107013>
13. Khader F, Mueller-Franzes G, Arasteh ST et al (2022) Medical diffusion: denoising diffusion probabilistic models for 3D medical image generation. <https://doi.org/10.48550/arXiv.2211.03364>
14. Kazerouni A, Aghdam EK, Heidari M et al (2023) Diffusion models in medical imaging: a comprehensive survey. *Med Image Anal* 88:102846. <https://doi.org/10.1016/j.media.2023.102846>
15. Zhang J, He X, Qing L et al (2022) BPGAN: brain PET synthesis from MRI using generative adversarial network for multi-modal Alzheimer's disease diagnosis. *Comput Methods Prog Biomed* 217:106676. <https://doi.org/10.1016/j.cmpb.2022.106676>
16. Xie T, Cao C, Cui Z et al (2024) Synthesizing PET images from high-field and ultra-high-field MR images using joint diffusion attention model. *Med Phys* 51:5250–5269. <https://doi.org/10.1002/mp.17254>
17. Banaie Boroujeni K, Sigona MK, Treuting RL et al (2022) Anterior cingulate cortex causally supports flexible learning under motivationally challenging and cognitively demanding conditions. *PLoS Biol* 20:e3001785. <https://doi.org/10.1371/journal.pbio.3001785>
18. Sudarshan VP, Upadhyay U, Egan GF et al (2021) Towards lower-dose PET using physics-based uncertainty-aware multimodal learning with robustness to out-of-distribution data. *Med Image Anal* 73:102187. <https://doi.org/10.1016/j.media.2021.102187>
19. Hussein R, Shin D, Zhao MY et al (2024) Turning brain MRI into diagnostic PET: 15O-water PET CBF synthesis from multi-contrast MRI via attention-based encoder–decoder networks. *Med Image Anal* 93:103072. <https://doi.org/10.1016/j.media.2023.103072>
20. Rajagopal A, Natsuaki Y, Wangerin K et al (2023) Synthetic PET via domain translation of 3-D MRI. *IEEE Trans Radiat Plasma Med Sci* 7:333–343. <https://doi.org/10.1109/TRPMS.2022.3223275>
21. Kocak B, Akinci D'Antonoli T, Mercaldo N et al (2024) METHodological RadiomicS Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMIL. *Insights Imaging* 15:8. <https://doi.org/10.1186/s13244-023-01572-w>
22. Vega F, Addeh A, Ganesh A et al (2024) Image translation for estimating two-dimensional axial amyloid-beta PET from structural MRI. *Magn Reson Imaging* 59:1021–1031. <https://doi.org/10.1002/jmri.29070>
23. Kim K, Byun BH, Lim I et al (2023) Deep learning-based delayed PET image synthesis from corresponding early scanned PET for dosimetry uptake estimation. *Diagnostics (Basel)* 13:3045. <https://doi.org/10.3390/diagnostics13193045>
24. Xue H, Zhang Q, Zou S et al (2021) LCPR-Net: low-count PET image reconstruction using the domain transform and cycle-consistent generative adversarial networks. *Quant Imaging Med Surg* 11:749–762. <https://doi.org/10.21037/qims-20-66>
25. Lei L, Zhang R, Zhang H et al (2023) Cross-modal PET synthesis method based on improved edge-aware generative adversarial network. *J Nanoelectron Optoelectron* 18:1184–1192. <https://doi.org/10.1166/jno.2023.3502>
26. Schaefferkoetter J, Yan J, Ortega C et al (2020) Convolutional neural networks for improving image quality with noisy PET data. *EJNMMI Res* 10:105. <https://doi.org/10.1186/s13550-020-00695-1>
27. Aghakhan Olia N, Kamali-Asl A, Hariri Tabrizi S et al (2022) Deep learning-based denoising of low-dose SPECT myocardial perfusion images: quantitative assessment and clinical performance. *Eur J Nucl Med Mol Imaging* 49:1508–1522. <https://doi.org/10.1007/s00259-021-05614-7>
28. Yuan S, Chen X, Liu Y et al (2023) Comprehensive evaluation of similarity between synthetic and real CT images for nasopharyngeal carcinoma. *Radiat Oncol* 18:182. <https://doi.org/10.1186/s13014-023-02349-7>
29. McNaughton J, Fernandez J, Holdsworth S et al (2023) Machine learning for medical image translation: a systematic review. *Bioengineering* 10:1078. <https://doi.org/10.3390/bioengineering10091078>
30. Fard AS, Reutens DC, Ramsay SC et al (2024) Image synthesis of interictal SPECT from MRI and PET using machine learning. *Front Neurol* 15:1383773. <https://doi.org/10.3389/fneur.2024.1383773>
31. Huang H, Yang G, Zhang W et al (2021) A deep multi-task learning framework for brain tumor segmentation. *Front Oncol* 11:690244. <https://doi.org/10.3389/fonc.2021.690244>
32. Sailunaz K, Bestepe D, Alhaji S et al (2023) Brain tumor detection and segmentation: interactive framework with a visual interface and feedback facility for dynamically improved accuracy and trust. *PLoS One* 18:e0284418. <https://doi.org/10.1371/journal.pone.0284418>
33. Jang S-I, Lois C, Thibault E et al (2023) TauPETGen: text-conditional tau PET image synthesis based on latent diffusion models. <https://doi.org/10.48550/arXiv.2306.11984>
34. Chen KT, Gong E, De Carvalho Macruz FB et al (2019) Ultra-low-dose ¹⁸F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. *Radiology* 290:649–656. <https://doi.org/10.1148/radiol.2018180940>
35. Miao T, Zhou B, Liu J et al (2023) Generation of whole-body FDG parametric K_i images from static PET images using deep learning. *IEEE Trans Radiat Plasma Med Sci* 7:465–472. <https://doi.org/10.1109/TRPMS.2023.3243576>
36. Zhou Y, Yang Z, Zhang H et al (2022) 3D segmentation guided style-based generative adversarial networks for PET synthesis. *IEEE Trans Med Imaging* 41:2092–2104. <https://doi.org/10.1109/TMI.2022.3156614>
37. Abazari MA, Soltani M, Moradi Kashkooli F, Raahemifar K (2022) Synthetic 18F-FDG PET image generation using a combination of biomathematical modeling and machine learning. *Cancers (Basel)* 14:2786. <https://doi.org/10.3390/cancers14112786>
38. Wang R, Liu H, Toyonaga T et al (2021) Generation of synthetic PET images of synaptic density and amyloid from ¹⁸F-FDG images using deep learning. *Med Phys* 48:5115–5129. <https://doi.org/10.1002/mp.15073>
39. Hussein R, Shin D, Zhao M et al (2022) Brain MRI-to-PET synthesis using 3D convolutional attention networks. <https://doi.org/10.48550/arXiv.2211.12082>
40. Lee J, Burkett BJ, Min H-K et al (2024) Synthesizing images of tau pathology from cross-modal neuroimaging using deep learning. *Brain* 147:980–995. <https://doi.org/10.1093/brain/awad346>
41. Mattoli MV, Giancipoli RG, Coccilillo F et al (2024) The role of PET imaging in patients with prion disease: a literature review. *Mol Imaging Biol* 26:195–212. <https://doi.org/10.1007/s11307-024-01895-0>
42. Wang Y-R, Baratto L, Hawk KE et al (2021) Artificial intelligence enables whole-body positron emission tomography scans with minimal radiation exposure. *Eur J Nucl Med Mol Imaging* 48:2771–2781. <https://doi.org/10.1007/s00259-021-05197-3>
43. 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. <https://doi.org/10.1016/j.media.2019.03.013>
44. Emami H, Liu Q, Dong M (2020) FREA-Unet: frequency-aware U-net for modality transfer. <https://doi.org/10.48550/arXiv.2012.15397>
45. Li J, Chen J, Tang Y et al (2023) Transforming medical imaging with Transformers? A comparative review of key properties, current progresses,

- and future perspectives. *Med Image Anal* 85:102762. <https://doi.org/10.1016/j.media.2023.102762>
46. Shamshad F, Khan S, Zamir SW et al (2022) Transformers in medical imaging: a survey. <https://doi.org/10.48550/arXiv.2201.09873>
 47. Wang Y, Zhou L, Yu B et al (2019) 3D Auto-context-based locality adaptive multi-modality GANs for PET synthesis. *IEEE Trans Med Imaging* 38:1328–1339. <https://doi.org/10.1109/TMI.2018.2884053>
 48. Wei W, Poirion E, Bodini B et al (2019) Predicting PET-derived demyelination from multimodal MRI using sketcher-refiner adversarial training for multiple sclerosis. *Med Image Anal* 58:101546. <https://doi.org/10.1016/j.media.2019.101546>
 49. Sanaat A, Arabi H, Mainta I et al (2020) Projection space implementation of deep learning-guided low-dose brain PET imaging improves performance over implementation in image space. *J Nucl Med* 61:1388–1396. <https://doi.org/10.2967/jnumed.119.239327>
 50. Shin H-C, Ihsani A, Mandava S et al (2020) GANBERT: generative adversarial networks with bidirectional encoder representations from transformers for MRI to PET synthesis. <https://doi.org/10.48550/arXiv.2008.04393>
 51. Islam J, Zhang Y (2020) GAN-based synthetic brain PET image generation. *Brain Inf* 7:3. <https://doi.org/10.1186/s40708-020-00104-2>
 52. Zhang S, Xu X, Li Q et al (2022) Brain network topology and structural-functional connectivity coupling mediate the association between gut microbiota and cognition. *Front Neurosci* 16:814477. <https://doi.org/10.3389/fnins.2022.814477>
 53. Lin W, Lin W, Chen G et al (2021) Bidirectional mapping of brain MRI and PET with 3D reversible GAN for the diagnosis of Alzheimer's disease. *Front Neurosci* 15:646013. <https://doi.org/10.3389/fnins.2021.646013>
 54. Sun H, Jiang Y, Yuan J et al (2022) High-quality PET image synthesis from ultra-low-dose PET/MRI using bi-task deep learning. *Quant Imaging Med Surg* 12:5326–5342. <https://doi.org/10.21037/qjms-22-116>
 55. Gao X, Shi F, Shen D, Liu M (2023) Multimodal transformer network for incomplete image generation and diagnosis of Alzheimer's disease. *Comput Med Imaging Graph* 110:102303. <https://doi.org/10.1016/j.compmedimag.2023.102303>
 56. Takita H, Matsumoto T, Tatekawa H et al (2023) AI-based virtual synthesis of methionine PET from contrast-enhanced MRI: development and external validation study. *Radiology* 308:e223016. <https://doi.org/10.1148/radiol.223016>
 57. Wang C, Piao S, Huang Z et al (2024) Joint learning framework of cross-modal synthesis and diagnosis for Alzheimer's disease by mining underlying shared modality information. *Med Image Anal* 91:103032. <https://doi.org/10.1016/j.media.2023.103032>
 58. Ouyang J, Chen KT, Duarte Armindo R et al (2024) Predicting FDG-PET images from multi-contrast MRI using deep learning in patients with brain neoplasms. *Magn Reson Imaging* 59:1010–1020. <https://doi.org/10.1002/jmri.28837>
 59. Sallam M, Barakat M, Sallam M (2024) A preliminary checklist (METRICS) to standardize the design and reporting of studies on generative artificial intelligence-based models in health care education and practice: development study involving a literature review. *Interact J Med Res* 13:e54704. <https://doi.org/10.2196/54704>

Publisher's Note

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