VM
VIA MEDICA

Performance of a deep learning enhancement method applied to PET images acquired with a reduced acquisition time

Krzysztof Ciborowski[1], Anna Gramek-Jedwabna[1], Monika Gołąb[1, 2], Izabela Miechowicz[3], Jolanta Szczurek[2], Marek Ruchała[2], Rafal Czepczyński[1, 2]

[1]Department of Nuclear Medicine, Affidea Polska, Poznan, Poland
[2]Department of Endocrinology, Metabolism and Internal Diseases, Poznan University of Medical Sciences, Poznan, Poland
[3]Department of Informatics and Statistics, Poznan University of Medical Sciences, Poznan, Poland

Abstract

Background: This study aims to evaluate the performance of a deep learning enhancement method in PET images reconstructed with a shorter acquisition time, and different reconstruction algorithms. The impact of the enhancement on clinical decisions was also assessed.

Material and methods: Thirty-seven subjects underwent clinical whole-body [^{18}F]FDG PET/CT exams with an acquisition time of 1.5 min per bed position. PET images were reconstructed with the OSEM algorithm using 66% counts (imitating 1 min/bed acquisition time) and 100% counts (1.5 min/bed). Images reconstructed from 66% counts were subsequently enhanced using the SubtlePET™ (SP) deep-learning-based software, (Subtle Medical, USA) — with two different software versions (SP1 and SP2). Additionally, images obtained with 66% counts were reconstructed with QClear™ (GE, USA) algorithm and enhanced with SP2. Volumes of interest (VOI) of the lesions and reference VOIs in the liver, brain, bladder, and mediastinum were drawn on OSEM images and copied on SP images. Quantitative SUV$_{max}$ values per VOI of OSEM or QClear™ and AI-enhanced 'shortened' acquisitions were compared.

Results: Two hundred and fifty-two VOIs were identified (37 for each reference region, and 104 for the lesions) for OSEM, SP1, SP2, and QClear™ images AI-enhanced with SP2. SUV$_{max}$ values on SP1 images were lower than standard OSEM, but on SP2 differences were smaller (average difference for SP1 11.6%, for SP2 -4.5%). For images reconstructed with QClear™, SUV$_{max}$ values were higher (average +8.9%, median 6.1%, SD 18.9%). For small lesions with SUV$_{max}$ values range 2.0 to 4.0 decrease of measured SUV$_{max}$ was much less significant with SP2 (for liver average -6.5%, median -5.6% for lesions average -5.6%, median — 6.0, SD 5.2%) and showed the best correlation with original OSEM. While no artifacts and good general diagnostic confidence were found in AI-enhanced images, SP1, the images were not equal to the original OSEM — some lesions were hard to spot. SP2 produced images with almost the same quality as the original 1.5 min/bed OSEM reconstruction.

Conclusions: The studied deep learning enhancement method can be used to accelerate PET acquisitions without compromising quantitative SUV$_{max}$ values. AI-based algorithms can enhance the image quality of accelerated PET acquisitions, enabling the dose reduction to the patients and improving the cost-effectiveness of PET/CT imaging.

KEY words: artificial intelligence; dose reduction; PET/CT; deep learning

Nucl Med Rev 2023; 26, 116–122

Correspondence to: Rafał Czepczyński, Department of Nuclear Medicine, Affidea Polska, 28 Czerwca 1956 St. 202, 61–485 Poznan, Poland; phone: +48 603758729; e-mail: rafal.czepczynski@affidea.com

Introduction

New algorithms based on artificial intelligence (AI), more precisely deep learning (DL), a branch of machine learning, designed for an improvement of medical imaging have been developed in recent years [1]. These novel methods, aimed to improve image quality, have been introduced also to positron emission tomography/computed tomography (PET/CT) — an imaging method that is widely used in many clinical applications, especially oncology, neurology, and cardiology [2, 3]. The long way from the idea of an image enhancement method through its validation up to the implementation of different DL algorithms in PET/CT into clinical use has been described elsewhere [4, 5].

Since the advent of the PET technique, a lot of solutions have been proposed to shorten the time of acquisition and, even more importantly, to reduce the radionuclide activity injected into the patients. Any change of either parameter always impacted image quality, quantitative data, and clinical confidence of the readers reporting the scans [6, 7]. The application of machine learning in nuclear medicine image analysis is very promising [8] and seems to be a perfect solution for the reduction of the administered radiopharmaceutical activity and, subsequently, patient radiation exposure without compromising the image quality and diagnostic performance of PET/CT [9]. That is why the interest in the application of AI-based methods is growing [10], with special attention to image enhancement [11] and improvement of image quality in low-dose acquisition [12].

An algorithm recently developed by Subtle Medical showed promising results in validation studies performed in Europe and on other continents [13, 14] and is now registered and commercially available. Its routine use in PET/CT imaging may potentially reduce the acquisition time or patient's radiation dose. Based on these encouraging results, our institution decided to use it in its 3 PET/CT centers. The purpose of this study is to determine, if the published early results can be replicated in daily clinical routine and if the DL image enhancement can be used to reduce the acquisition time or the injected radiotracer activity.

Material and methods

Subsequent, randomly assigned patients undergoing a routine PET/CT with fluorine-18-deoxyglucose ([18F]FDG) for oncological indications were qualified for the study. After the initial evaluation, images with no apparent [18F]FDG-positive lesion (normal PET/CT scans) were disqualified. As a result, only patients with at least one [18F]FDG-positive, pathological finding, were qualified.

All patients were injected with the standard dose of 3.5 MBq/kg [18F]FDG and the images were acquired approximately 60 min post-injection. The images were obtained with the GE Discovery IQ PET/CT scanner (GE Healthcare, Milwaukee, WI, USA). The scanner is using the standard OSEM reconstruction, as well as the QClear™ enhanced lesion detection reconstruction [15, 16]. Patients underwent a low-dose computed tomography (CT) prior to PET-processed acquisition, for attenuation correction and anatomical correlation of PET findings. The whole-body PET/CT exams were performed with a standard acquisition time of 1.5 min per bed position. Emission data was corrected for randoms, dead time, scatter, and attenuation and was reconstructed iteratively

by an ordered-subsets expectation maximization (OSEM) algorithm. The original OSEM reconstructed images (100% counts, 1.5 min/bed) were sent from the modality to the SubtleEdge server for processing using deep learning (DL) based software (SubtlePET™, Subtle Medical, Menlo Park, CA, USA). SubtlePET™ reconstruction software was developed using 2.5D encoder-decoder U-Net with the main purpose of denoising the images. It employs a convolutional neural network (CNN)-based method in a pixel's neighborhood to reduce noise and increase image quality [17, 18].

To simulate images obtained with a shorter acquisition time or lower [18F]FDG activity, the 1 min/bed (equivalent to 66% of acquired counts or 2.33 MBq/kg of injected activity) were also sent to the DL software, for comparison. As the manufacturer offered two different versions of the software, both versions were used for processing: the original SubtlePET version 1 (SP1), and an upgraded SubtlePET version 2 (SP2).

Additionally, to test the performance of the DL processing in case of a reconstruction algorithm other than OSEM, the 1 min/bed images reconstructed with the QClear™ (GE, Milwaukee, USA) reconstruction algorithm were also sent for the processing, however with the version 2 of the software (SP2) only. To summarize, the following datasets were obtained and analyzed:
— 1.5 min/bed OSEM (the original PET images)
— 1 min/bed SubtlePET™ v1 over OSEM (SP1/OSEM)
— 1 min/bed SubtlePET™ v2 over OSEM (SP2/OSEM)
— 1 min/bed SubtlePET™ v2 over QClear™ (SP2/QClear).

An experienced, board-certified nuclear medicine physician reviewed the standard acquisition PET images in the Advantage Workstation (GE, Milwaukee, WI, USA) and identified possible lesions. 30 mm radius spherical volumes of interest (VOIs) of the lesions and reference VOIs placed in the reference organs: in the liver, brain, bladder, and mediastinum were drawn on the standard acquisition images. The same lesions, VOIs, and reference VOIs were subsequently copied and reviewed on the AI-enhanced images (SP1/OSEM, SP2/OSEM, and SP2/QClear). For each VOI, SUV_{max} values normalized to lean body mass were calculated at each type of reconstruction.

Differences in values between the standard study (1.5 min/bed OSEM) and copy of the VOIs for additional datasets, both, in SUV_{max} and as percentage (%) were statistically evaluated: average difference, median and standard deviation for each dataset, reference region and all the lesions together were compared. In order to test the diagnostic performance of AI-enhancement in the case of lesions with relatively low [18F]FDG uptake, a similar analysis of a subgroup of scans presenting lesions with $SUV_{max} < 4.0$ was performed separately.

Statistical analysis

Statistical analysis was performed for two ranges of results: 1) all measured lesions [with four methods, i.e. 1.5 min/bed OSEM and three methods for 1 min/bed acquisition time: SP1/OSEM, SP2/OSEM, and SP2/QClear) and 2) for lesions with low uptake ($SUV_{max} < 4$) also with four mentioned methods. Descriptive statistics for categorical variables were presented as relative/absolute frequencies, while those for continuous ones as the median (range).

Table 1. Population characteristic

Patient characteristic	Number [%] (N = 37)
Female	16 (43.2%)
Male	21 (56.8%)
	Median (range)
Uptake time	59 min (49–74)
Body height	165 cm (151–190)
Body weight	73 kg (36–109)
Body mass index	25.2 kg/m^2 (15.2–37.3)
Age	67 years (16–80)
Tumor characteristic	Number [%] (N = 37)
Hodgkin/non-Hodgkin lymphoma	13 (35.1%)
Lung cancer	11 (29.7%)
Colorectal cancer	5 (13.5%)
Esophageal cancer	3 (8.1%)
Breast cancer	3 (8.1%)
Laryngeal cancer	2 (5.4%)

As all measured values had a non-normal distribution, we performed the Friedman test with the calculation of the Kendall correlation coefficient and later Spearman rank correlation for correlation between the base dataset (OSEM 1.5 min/bed) and datasets reconstructed with reduced time.

For lesions with $SUV_{max} < 4$, a normal distribution was found, but without preserved sphericity. Therefore, the Friedman test was performed, followed by the r-Pearson linear correlation coefficient calculation. Next ICC intraclass correlation was calculated between OSEM 1.5 min/bed and each of the AI datasets. ICC was interpreted according to the Landis interpretation scale [19] (0.0: poor; 0.0–0.20: slight; 0.21–0.40: fair; 0.41–0.60: moderate; 0.61–0.80: substantial; 0.81–1.00: almost-perfect reproducibility).

Results

Thirty-seven subjects (21 males, 16 females, aged 16–80 years; median age 67 years) with at least one [18F]FDG-avid lesion were qualified for the study. Patients' weight varied from 36 to 109 kg (median 73 kg) and body mass index (BMI) ranged from 15.1 to 37.2 (median 25.2). Patient characteristics and clinical indications for the [18F]FDG PET/CT imaging are shown in Table 1.

We identified altogether 252 VOIs: 104 lesion VOIs (at least one lesion, no more than four lesions per patient) and 148 reference VOIs (37 for each of four reference regions — liver, brain, bladder, and mediastinum). SUV_{max} values obtained with all the image reconstructions and enhancement methods were compared with the SUV_{max} values measured at the original images (1.5 min/bed OSEM). The differences in SUV_{max} values in absolute numbers and percentage are presented in Table 2. SUV_{max} values on AI-enhanced images were lower than on unenhanced standard OSEM images but using SP2 the difference was smaller (median difference for SP1 was −11.89%, for SP2 −4.82%; Tab. 2). Still, all new reconstruction methods showed a strong positive correlation to the original OSEM 1.5 min/bed data. For the images reconstructed with QClear™, the trend was reverse — SUV_{max} values were higher than on unenhanced OSEM images (median 6.12%). For lesions with SUV_{max} values below 4.0 — the decrease of measured SUV_{max} was much less significant with SP2 (for liver reference median −5.6%, for lesions median −6.04%). Statistical analysis showed no difference between OSEM 1.5, SP2/OSEM, and SP2/Qclear, with the highest Pearson r and ICC intraclass coefficients for data reconstructed with SP2 (Tab. 3).

Statistical analysis showed that for lesions with SUV_{max} range 2.0 to 4.0, the SP2 version strongly correlates with OSEM; r-Person correlation coefficient is 0.9166 and ICC intraclass correlation of 0.9545 (almost perfect reproducibility), it is also clearly visible on Bland–Altman plot (Fig. 1).

Also for lesions with $SUV_{max} < 4$, SP2/OSEM results showed the best correlation with the original OSEM data (Fig. 2).

In the qualitative evaluation, generally, good image quality was found in SP1/OSEM, with no apparent artifacts. The images were not equal to the original OSEM ones — some lesions found in mediastinum were hard to detect but they were still identifiable. SP2/OSEM provided more detailed images (less smoothing effect) and almost the same quality as the original 1.5 min/bed standard reconstruction (much better than 1 min/bed OSEM reconstruction) (Fig. 3).

Discussion

The study was performed in parallel with the implementation of SubtlePET™ reconstruction in clinical routine, based on a previous assessment performed by Katsari et al. [13], and performed in a similar timeframe, already published by Weyts et al. [17] and Bonardel et al. [18]. The main goal was to reduce the dose applied to the patient (by 1/3), with no negative effect on the image quality and clinical confidence. Reduction of the [18F]FDG injected activity could also translate to beneficial cost-effectiveness of PET/CT

Table 2. SUV_{max} differences (values and %) between 1.5 min OSEM and 1 min SP1/OSEM, SP2/OSEM, SP2/QClear

1 min/bed		Reference regions				Lesions	
		Liver	Brain	Bladder	Mediastinum	$SUV_{max} < 4$	All SUV_{max} values
SP1/OSEM	Median	−0.15	−0.17	0.07	−0.19	−0.61	−0.50
		(−6.49%)	(−2.04%)	(0.58%)	(−9.58%)	(−11.51%)	(−11.89%)
SP2/OSEM	Median	−0.14	−0.17	−0.15	−0.10	−0.31	−0.25
		(−5.60%)	(−2.03%)	(−0.86%)	(−5.28%)	(−6.04%)	(−4.82%)
SP2/QClear	Median	−0.45	−0.15	0.95	−0.29	0.63	0.38
		(−16.55%)	(−1.81%)	(3.62%)	(−14.29%)	(11.31%)	(6.12%)

Krzysztof Ciborowski et al., Deep learning in PET

Table 3. Results correlation between 1.5 min OSEM and 1 min SP1/OSEM, SP2/OSEM, SP2/QClear

		Lesions	
		$SUV_{max} < 4$ (normal distribution)	All SUV_{max} values (inconsistent with normal distribution)
SP1/OSEM	r-Person correlation coefficient	0.9058	–
	ICC intraclass correlation	0.9498	–
	Rank-order correlation R Spearman	–	0.9863
	Kendall coefficient	–	0.9931
SP2/OSEM	r-Person correlation coefficient	0.9166	–
	ICC intraclass correlation	0.9545	–
	Rank-order correlation R Spearman	–	0.9913
	Kendall coefficient	–	0.9956
SP2/QClear	r-Person correlation coefficient	0.8459	–
	ICC intraclass correlation	0.8746	–
	Rank-order correlation R Spearman	–	0.9710
	Kendall coefficient	–	0.9855
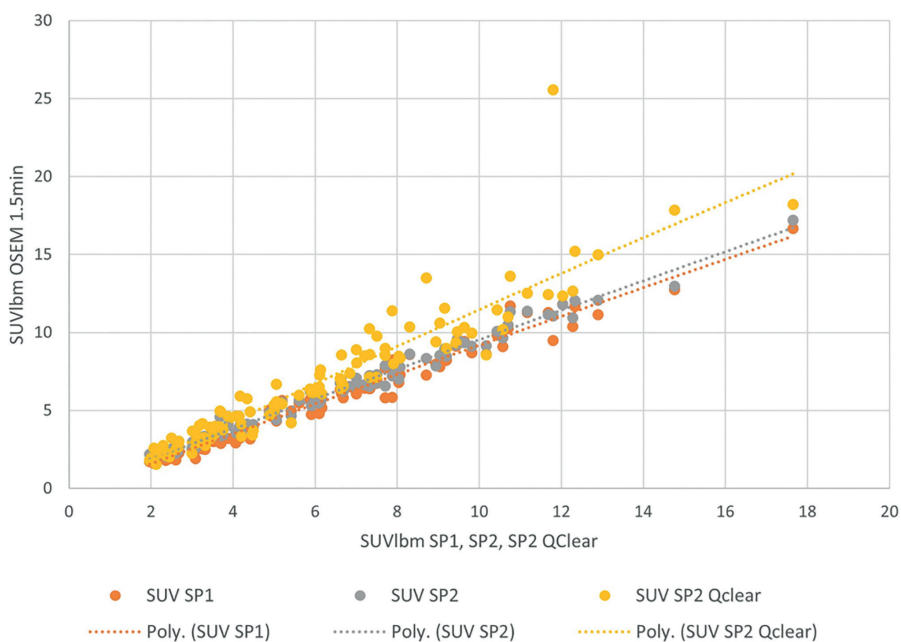
Figure 1. Bland Altman plot 1.5 min OSEM — 1 min SP2/OSEM

procedure. In contrast to very encouraging published data, we faced some negative feedback from physicians reporting PET/CT examinations. The physicians mostly complained of the low visibility of small changes and lower SUV_{max} values, that, in some cases, had an impact on the clinical interpretation of the lesions. We decided to perform quantitative analyses using normal dose (time) and to simulate lower dose retrospectively by reconstructing the images with the time per bed shorter by 1/3, i.e. 1 min/bed instead of 1.5 min/bed.

The initial results obtained with version 1 of the software (SP1/OSEM) confirmed some of the observed problems, such as excessive smoothing of the images, low SUV_{max} value, and a potential lower detectability of small [18F]FDG-avid lesions. Once the manufacturer developed a new version of the algorithm (SP2), addressing most of the raised issues, we decided to reevaluate the data with the new version.

We were able to reprocess previously acquired data with the new algorithm, applying it also to 1 min/bed data reconstructed with QClear™ reconstruction that is routinely used in our center, and to repeat all the measurements for all four types of data.

After a direct comparison of original clinical data and the reduced time DL algorithm outputs, the comparison of multiple
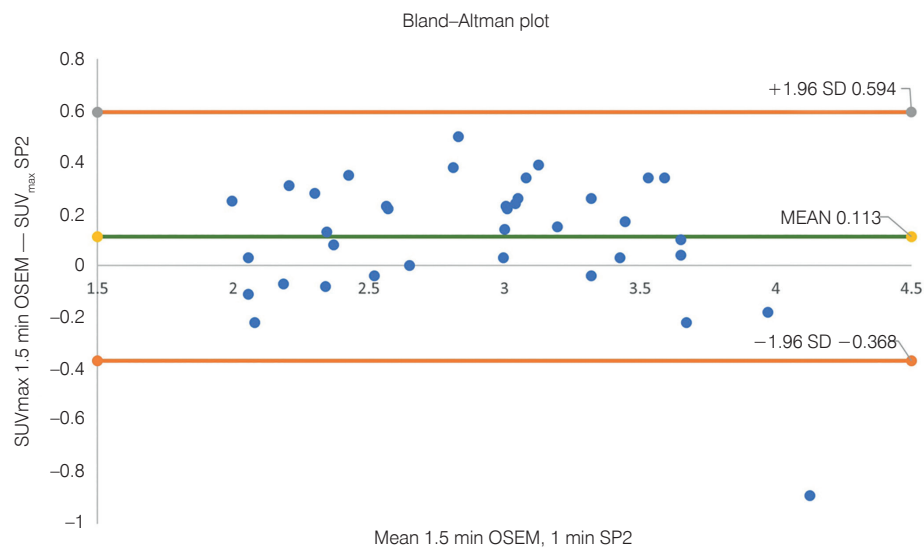
Figure 2. SUV$_{max}$ difference from 1.5 min OSEM for lesions with SUV$_{max}$ < 4 for SP1 (blue dots), SP2 (orange dots), and SP2 QClear (grey dots)
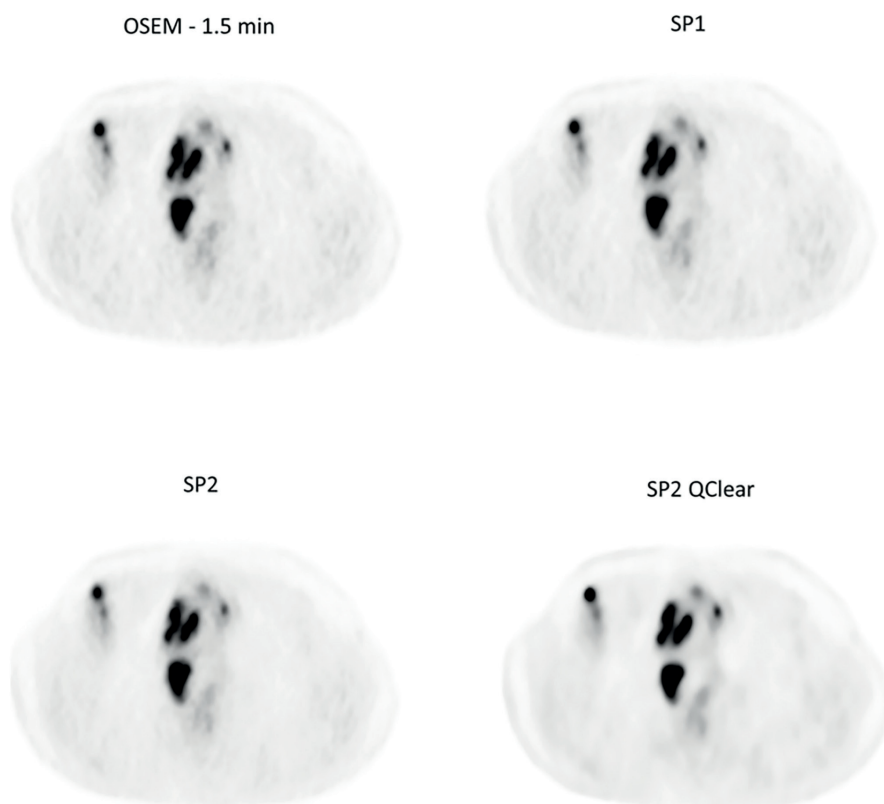

Figure 3. [18F]FDG PET/CT maximum intensity projection image, comparison of axial images — 1.5 min OSEM; 1 min — SP1/OSEM; SP2/OSEM; SP2/QClear. The example image represents an axial slice of the thorax in a patient with Hodgkin lymphoma. [18F]FDG-avid positive lesions were found in the mediastinum and right axillary region

outputs was conducted. We could see not only the impact of the AI-based algorithm on the lower count/dose or time, but also the differences between the two versions of the algorithm, and the impact of the reconstruction algorithm used on the final image.

It has to be pointed out that the DL algorithm is applied to already reconstructed and corrected (for attenuation, scatter, well counter, etc.) images, working on already statistically modified images and not on the original raw data.

The final number of patients in the cohort was a little lower than originally planned but still, it was large enough to provide a good representation of different clinical diagnoses, age, sex, and body mass index of the subjects.

The main limitation of the study is its one center/one PET/CT scanner setting. We are planning to perform a larger study on data obtained from 3–4 scanners in different centers. Another limitation is related to the comparison method as only one dose of [^{18}F]FDG could be administered in each patient due to the radiation protection restrictions. In contrast to the study by Katsari et al. [13], and similarly to Weyts et al. [17] and Bonardel et al. [18], we decided to keep standard dose and acquisition time and to reconstruct the images retrospectively with a shorter time (1/3 reduction) to simulate lower dose, rather than to administer lower activities and acquire for a longer time to simulate normal dose. It gave us also the possibility to check if the subjects had any lesions visible on the standard OSEM images before reprocessing images with the DL algorithm, in order to ascertain the presence of lesions feasible for the evaluation.

It should be also pointed out how the DL algorithm in question was qualified by us as users. While in other modalities (like CT or MR), AI provides already some algorithms supporting the clinical decisions of the readers, the SubtlePET algorithm is still just an image enhancement tool with no support for the clinical interpretation.

We also need to remember that we were working on the already approved and registered algorithm, while many other centers try to develop and test their own methods [11, 12, 20].

Conclusions

The evaluated AI-based image enhancement can be used to accelerate PET acquisitions by one-third without compromising quantitative SUV$_{max}$ values and image quality as compared to the acquisitions with standard duration. Respectively, it can also be used to optimize PET images obtained with a radiopharmaceutical activity reduced by one-third. Subsequently, routine use of the image enhancement may increase PET/CT scanner throughput or reduce the patient's radiation burden. Furthermore, the analyzed AI enhancement can be used in PET images processed with different reconstruction algorithms. However, as shown by our experience with different AI algorithms, one should be aware of its impact on SUV$_{max}$ values and lesion detectability.

Conflict of interests

Authors do not claim any conflict of interests.

Funding

No funding was necessary.

References

1. Nensa F, Demircioglu A, Rischpler C. Artificial intelligence in nuclear medicine. J Nucl Med. 2019; 60(Suppl 2): 29S–37S, doi: 10.2967/jnumed.118.220590, indexed in Pubmed: 31481587.

2. Salem AE, Shah HR, Covington MF, et al. PET-CT in clinical adult oncology: I. hematologic malignancies. Cancers (Basel). 2022; 14(23), doi: 10.3390/cancers14235941, indexed in Pubmed: 36497423.

3. Zhao HB, Liu C, Ye J, et al. A comparison between deep learning convolutional neural networks and radiologists in the differentiation of benign and malignant thyroid nodules on CT images. Endokrynol Pol. 2021; 72(3): 217–225, doi: 10.5603/EP.a2021.0015, indexed in Pubmed: 33619712.

4. Arabi H, Zaidi H. Applications of artificial intelligence and deep learning in molecular imaging and radiotherapy. Eur J Hybrid Imaging. 2020; 4(1): 17, doi: 10.1186/s41824-020-00086-8, indexed in Pubmed: 34191161.

5. Sollini M, Bartoli F, Marciano A, et al. Artificial intelligence and hybrid imaging: the best match for personalized medicine in oncology. Eur J Hybrid Imaging. 2020; 4(1): 24, doi: 10.1186/s41824-020-00094-8, indexed in Pubmed: 34191197.

6. Wyszomirska A, Czepczyński R, Sałyga A, et al. Effect of acquisition time on the measurement of SUV parameter. Eur J Nucl Med Mol Imaging. 2012; 39: S508–S509.

7. Aide N, Lasnon C, Kesner A, et al. New PET technologies — embracing progress and pushing the limits. Eur J Nucl Med Mol Imaging. 2021; 48(9): 2711–2726, doi: 10.1007/s00259-021-05390-4, indexed in Pubmed: 34081153.

8. Arabi H, AkhavanAllaf A, Sanaat A, et al. The promise of artificial intelligence and deep learning in PET and SPECT imaging. Phys Med. 2021; 83: 122–137, doi: 10.1016/j.ejmp.2021.03.008, indexed in Pubmed: 33765602.

9. Reader AJ, Schramm G. Artificial intelligence for PET image reconstruction. J Nucl Med. 2021; 62(10): 1330–1333, doi: 10.2967/jnumed.121.262303, indexed in Pubmed: 34244357.

10. Visvikis D, Lambin P, Beuschau Mauridsen K, et al. Application of artificial intelligence in nuclear medicine and molecular imaging: a review of current status and future perspectives for clinical translation. Eur J Nucl Med Mol Imaging. 2022; 49(13): 4452–4463, doi: 10.1007/s00259-022-05891-w, indexed in Pubmed: 35809090.

11. Mehranian A, Wollenweber SD, Walker MD, et al. Deep learning-based time-of-flight (ToF) image enhancement of non-ToF PET scans. Eur J Nucl Med Mol Imaging. 2022; 49(11): 3740–3749, doi: 10.1007/s00259-022-05824-7, indexed in Pubmed: 35507059.

12. Xue S, Guo R, Bohn KP, et al. A cross-scanner and cross-tracer deep learning method for the recovery of standard-dose imaging quality from low-dose PET. Eur J Nucl Med Mol Imaging. 2022; 49(6): 1843–1856, doi: 10.1007/s00259-021-05644-1, indexed in Pubmed: 34950968.

13. Katsari K, Penna D, Arena V, et al. Artificial intelligence for reduced dose 18F-FDG PET examinations: a real-world deployment through a standardized framework and business case assessment. EJNMMI Phys. 2021; 8(1): 25, doi: 10.1186/s40658-021-00374-7, indexed in Pubmed: 33687602.

14. Xu F, Pan B, Zhu X, et al. Evaluation of deep learning based PET image enhancement method in diagnosis of lymphoma. J Nucl Med. 2020; 61(Supp.1): 431.

15. Reynés-Llompart G, Gámez-Cenzano C, Romero-Zayas I, et al. Performance characteristics of the whole-body discovery IQ PET/CT system. J Nucl Med. 2017; 58(7): 1155–1161, doi: 10.2967/jnumed.116.185561, indexed in Pubmed: 28302761.

16. Wyrzykowski M, Siminiak N, Kaźmierczak M, et al. Impact of the Q.Clear reconstruction algorithm on the interpretation of PET/CT images in patients with lymphoma. EJNMMI Res. 2020; 10(1): 99, doi: 10.1186/s13550-020-00690-6, indexed in Pubmed: 32845406.

17. Weyts K, Lasnon C, Ciappuccini R, et al. Artificial intelligence-based PET denoising could allow a two-fold reduction in [F]FDG PET acquisition time in digital PET/CT. Eur J Nucl Med Mol Imaging. 2022; 49(11): 3750–3760, doi: 10.1007/s00259-022-05800-1, indexed in Pubmed: 35593925.

18. Bonardel G, Dupont A, Decazes P, et al. Clinical and phantom validation of a deep learning based denoising algorithm for F-18-FDG PET images from lower detection counting in comparison with the standard acquisition. EJNMMI Phys. 2022; 9(1): 36, doi: 10.1186/s40658-022-00465-z, indexed in Pubmed: 35543894.

19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1): 159–174, indexed in Pubmed: 843571.

20. Sanaat A, Shiri I, Arabi H, et al. Deep learning-assisted ultra-fast/low-dose whole-body PET/CT imaging. Eur J Nucl Med Mol Imaging. 2021; 48(8): 2405–2415, doi: 10.1007/s00259-020-05167-1, indexed in Pubmed: 33495927.