Journal of Nuclear Medicine, published on January 10, 2020 as doi:10.2967/jnumed.119.239327

Projection-space implementation of deep learning-guided low-dose brain PET imaging improves performance over implementation in image-space

Amirhossein Sanaat[1], Hossein Arabi[1], Ismini Mainta[1], Valentina Garibotto[1,2] and Habib Zaidi[1,2,3,4]

[1]Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva CH-1211, Switzerland
[2]Geneva University Neurocenter, Geneva University, 1205 Geneva, Switzerland
[3]Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands
[4]Department of Nuclear Medicine, University of Southern Denmark, DK-500, Odense, Denmark

Corresponding author:

Habib Zaidi, Ph.D
Geneva University Hospital
Division of Nuclear Medicine and Molecular Imaging
CH-1211 Geneva, Switzerland
Tel: +41 22 372 7258
Fax: +41 22 372 7169
email: habib.zaidi@hcuge.ch

First Author:

Amirhossein Sanaat
Geneva University Hospital
Division of Nuclear Medicine and Molecular Imaging
CH-1211 Geneva, Switzerland
Tel: +41 778167119
email: amirhossein.sanaat@etu.unige.ch

Short running title: Deep learning for PET image denoising

ABSTRACT

Purpose: To assess the performance of full dose (FD) positron emission tomography (PET) image synthesis in both image and projection space from low-dose (LD) PET images/sinograms without sacrificing diagnostic quality using deep learning techniques.

Methods: Clinical brain PET/CT studies of 140 patients were retrospectively employed for LD to FD PET conversion. 5% of the events were randomly selected from the FD list-mode PET data to simulate a realistic LD acquisition. A modified 3D U-Net model was implemented to predict FD sinograms in the projection-space (PSS) and FD images in image-space (PIS) from their corresponding LD sinograms/images, respectively. The quality of the predicted PET images was assessed by two nuclear medicine specialists using a five-point grading scheme. Quantitative analysis using established metrics including the peak signal-to-noise ratio (PSNR), structural similarity index metric (SSIM), region-wise standardized uptake value (SUV) bias, as well as first-, second- and high-order texture radiomic features in 83 brain regions for the test and evaluation dataset was also performed.

Results: All PSS images were scored 4 or higher (good to excellent) by the nuclear medicine specialists. PSNR and SSIM values of 0.96 ± 0.03, 0.97 ± 0.02 and 31.70 ± 0.75, 37.30 ± 0.71 were obtained for PIS and PSS, respectively. The average SUV bias calculated over all brain regions was 0.24 ± 0.96% and 1.05 ± 1.44% for PSS and PIS, respectively. The Bland-Altman plots reported the lowest SUV bias (0.02) and variance (95% CI: -0.92, +0.84) for PSS compared with the reference FD images. The relative error of the homogeneity radiomic feature belonging to the Grey Level Co-occurrence Matrix category was -1.07 ± 1.77 and 0.28 ± 1.4 for PIS and PSS, respectively

Conclusion: The qualitative assessment and quantitative analysis demonstrated that the FD PET prediction in projection space led to superior performance, resulting in higher image quality and lower SUV bias and variance compared to FD PET prediction in the image domain.

Keywords: PET/CT, brain imaging, low-dose imaging, deep learning, radiomics.

INTRODUCTION

Molecular neuroimaging using is PET is ideally suited for monitoring cell/molecular events early in the course of a neurodegenerative disease as well as during pharmacologic therapy (*1*). PET is a molecular imaging technique that produces a three-dimensional (3D) radiotracer distribution map representing properties of biological tissues, such as metabolic activity or receptor availability. PET images suffer from relatively high noise level dictated by the Poisson nature of annihilation photons emission and detection. Apart from the technical aspects, PET image quality depends on the amount of injected radiotracer and/or acquisition time which are proportional to the statistics of the detected events. The main argument in favour of a reduction of the injected radiotracer's activity is linked to the potential risks of ionizing radiation (*2*). Albeit low, this risk increase motivates precaution, particularly in paediatric patients, healthy volunteers or in case of multiple scanning for follow-up or monitoring the response to treatment using different tracers. Therefore, there has always been a desire to moderate the injected activity to minimize the potential health hazards. A reduced acquisition time could have a positive impact on patient's comfort and on scanner's throughput. However, dose/time reduction can adversely impact image quality, inevitably leading to lower signal-to-noise ratio (SNR), thus hampering the quantitative and diagnostic value of PET imaging.

To address this issue, a number of approaches have been proposed in the literature to produce standard/full-dose (FD) PET images from corresponding low-dose/count (LD) images (*3*). Formerly, iterative reconstruction algorithms with accurate statistical modelling (*4*) and post-processing/filtering (*5,6*) were the two common methods. However, these approaches tend to reduce spatial resolution, quantitative accuracy by producing over-smooth structures (*7,8*). In the past years, deep learning algorithms have witnessed notable growth in the fields of computer vision and medical image analysis (*9,10*). Contrary to other denoising approaches which are applied directly on LD PET images, deep learning algorithms are capable of learning a non-linear transformation to predict standard-dose images from low-dose inputs. In particular, convolutional neural network (CNN) models have demonstrated outstanding performance in cross-modality image synthesis, such as MRI to CT conversion (*11,12*), joint PET attenuation and scatter correction in image-space (*13,14*) as well as synthesis of FD PET images from LD images (*15-21*). Xiang *et al.* suggested a deep auto-context CNN architecture that estimates FD PET images based on local patches in LD PET images (*19*). A major limitation of this work is that 2D transaxial slices were extracted from PET images and used for 2D training of the CNN model. Another group claimed that reliable FD PET could be estimated from a 200th LD images using a residual U-Net architecture (*22*). Other work from the same group utilized 2D slices of LD ^{18}F-Florbetaben PET images along with various MR sequences, such as T1, T2 and DWI, to predict FD images using a U-Net architecture (*15*). Häggström *et al.* (*23*) developed a deep encoder-decoder network for direct reconstruction of PET images from sinograms whereas Hong *et al.* (*24*)

proposed a data-driven, single-image super-resolution technique for sinograms using a deep residual CNN to improve PET's spatial resolution and noise properties.

A more recent work reported on the use of a 3D-Unet along with anatomical information from co-registered MRI to improve PET's SNR without using higher SNR PET images in the training dataset (*25*). Cui *et al.* (*26*) presented an unsupervised model for PET denoising where the model was fed by the patient's prior high quality images and the noisy PET image itself was used as the training label. As such, this approach does not need any paired dataset for training. Furthermore, Lu *et al.* (*27*) investigated the effect of different network architectures and other parameters pertaining to both noise reduction and quantitative performance. The optimized fully 3D U-Net architecture is capable of reducing the noise in LD PET images while minimizing the quantification bias for lung nodules characterization.

Previous studies relied on deep learning-based approaches to establish an end-to-end pipeline to synthesize FD PET in image-space (*15-21*). As such, these approaches are optimized for a specific protocol, such as image reconstruction algorithm or post-reconstruction filter. Therefore, adoption to a different reconstruction technique requires retraining the CNN. Conversely, the prediction of FD PET images in the projection-space allows the selection of any reconstruction and/or post-reconstruction filter without the need for retraining the CNN. Furthermore, projection-space provides more comprehensive data representation than image-space, effectively containing detailed information about count statistics and spatial and temporal distributions. To take advantage of this fact, a 3D-Unet was trained to predict a FD sinograms from LD sinograms in an end-to-end fashion. Thereafter, the synthesized sinogram can be reconstructed using any reconstruction algorithm. The results achieved using the proposed framework are compared to the image domain implementation using the same 3D-Unet architecture. Low-dose PET imaging using the proposed approach would be beneficial in pediatric and adolescent clinical studies as well as in research protocols requiring serial studies.

MATERIALS AND METHODS

PET/CT Data Acquisition

The present study was conducted on ^{18}F-FDG brain PET/CT studies collected between June 2017 and May 2019 at Geneva university Hospital. The database consisted of 140 patients presenting with cognitive symptoms of possible neurodegenerative disease (73 ± 8 yrs), 66 males and 74 females (73 ± 9 yrs and 72 ± 11 yrs, respectively). Detailed demographic information of the patients is summarized in Table 1. The study protocol was approved by the institution's ethics committee and all patients gave written informed content. The PET/CT acquisitions were performed on a Biograph mCT scanner (Siemens Healthcare, Erlangen, Germany) about 35 minutes post-injection. A low-dose CT scan (120 kVp, 20 mAs) was performed for PET attenuation correction. The patients underwent a 20-min brain PET/CT scan after

injection of 205 ± 10 MBq of ^{18}F-FDG. PET data were acquired in list-mode format and reconstructed using e7 tool (an offline reconstruction toolkit provided by Siemens Healthcare) to produce FD PET sinograms/images. Subsequently, a subset of PET data containing 5% of the total events was extracted randomly from the list-mode data to produce LD sinograms (400×168×621 matrix) using a validated code (*28*). Both FD and LD PET images were reconstructed into a 200×200×109 image matrix (2.03×2.03×2.2 mm^3 voxel size) using an ordinary Poisson ordered subsets-expectation maximization (OP-OSEM) algorithm (5 iterations, 21 subsets). PET images underwent post-reconstruction Gaussian filtering with 2 mm FWHM similar to the clinical protocol.

U-Net Architecture

A modified 3D U-Net based on the model proposed in (*29*) was developed to predict FD images/sinograms (PIS/PSS) from their corresponding LD images/sinograms. Figure 1 shows the structure of the modified 3D U-Net, which consists of an encoder-decoder module. In the encoder part, each layer contains two 3D convolutions (*30*) followed by a rectified linear unit (ReLu) activation function and a 3D maxpooling with stride size of 2. In the decoder part, each layer consists of 3D up-sampling with stride of 2 followed by two 3D convolutions and a ReLu. The size of all convolutional kernels is 3×3 ×3 voxels in each convolutional layer. The shortcut connections between the outputs of each layer in the encoder network and the corresponding layer in the decoder network aimed at addressing the gradient vanishing problem that occurs in complex deep learning models. In CNN, the bottleneck is a layer which contains less neurons compared to its neighboring layers (*31*). To avoid this issue, the number of channels was doubled before maxpooling and before each ReLu function. The networks input are either a 101 × 101 × 71 matrix (after cropping) in the image domain or 400×168×62 matrix in the projection (sinogram) domain.

The modified 3D U-net architecture also includes a series of pooling options, dilated convolutional layers and 16 convolutional layers. The Adam optimizer with a learning rate of 0.001 was used to minimize the loss function. A dataset of paired LD and FD images/sinograms of 100 subjects were used to train the network using the adaptive moment estimation implemented in Keras open-source package (*32,33*), which computes adaptive learning rates for each parameter and saves an exponentially decaying average of past gradients using Eqs. 1 and 2:

$$L_t = r_1 L_{t-1} + (1 - r_1) g_t \quad (1)$$

$$V_t = r_2 V_{t-1} + (1 - r_2) g_t^2 \quad (2)$$

where L_t and V_t indicate the estimation of the mean and the uncentered variance of the gradients, respectively. g_t denotes the gradient at subsequent time step t and r_1 and r_2 are exponential decay rates with $r_1, r_2 \in [0, 1)$.

The model was implemented on NVIDIA 2080Ti GPU with 8 GB memory running under windows 10 operating system. The training was performed using mini-batch size of 5 for 250 epochs.

Data Augmentation. To increase the size of the training dataset while avoiding overfitting, three types of data augmentation methods were implemented. This included rotations, transformations, and zooming, which were randomly applied to the training data set. Hence, the model was trained using the 300 augmented along with the 100 original images. Applying such a rigid deformation to the training dataset assisted the network to learn features that are invariant to these transformations (*34*).

Training, Validation and Testing. The training and hyper parameter fine-tuning of the model were performed on 100 patients. Twenty patients were used for model evaluation whereas a separate unseen dataset of 20 patients served as test dataset. The mean squared error (MSE) loss function was used for the training of the model.

Evaluation Strategy

Clinical Qualitative Assessment. The predicted PSS and PIS FD PET images along with their corresponding reference FD and LD PET images were anonymized and randomly enumerated for qualitative evaluation by two nuclear medicine physicians. In total, 80 PET images were evaluated, including 20 reference FD, 20 LD, 20 PIS and 20 PSS PET images belonging to the test data set. The quality of PET images was assessed using a five-point grading scheme, namely 1: uninterpretable, 2: poor, 3: adequate, 4: good and 5: excellent (see Supplemental Figure 1).

Quantitative Analysis. The accuracy of the predicted FD images from LD PET data were evaluated using three quantitative metrics, including the root mean squared error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity index metrics (SSIM) (Eqs. 3-5, respectively). Moreover, these metrics were also calculated for the LD images to provide an insight about the noise levels and significant signal in the LD images.

$$RMSE(X,Y) = \sqrt{\frac{\sum_{j=1}^{L}(X-Y)^2}{L}} \tag{3}$$

$$PSNR(X,Y) = 20 \times log_{10}^{\left(\frac{Max(Y)}{\sqrt{MSE(X,Y)}}\right)} \tag{4}$$

$$SSIM(X,Y) = \frac{(2m_x m_y + c_1)(2\sigma_{xy} + c_2)}{(m_x^2 + m_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{5}$$

In Eq. (3), L is the total number of voxels in the head region, X is the reference image (FD), and Y is the predicted FD image. In Eq. (4) $Max(Y)$ indicates the maximum intensity value of X or Y, whereas *MSE* is the mean squared error. m_x and m_y in Eq. (5) denote the mean value of the images X and Y, respectively. σ_{xy} indicates the covariance of σ_x and σ_y, which in turn represent the variances of X and Y images, respectively. The constant parameters c_1 and c_2 ($c_1 = 0.01$ and $c_2 = 0.02$) were used to avoid a division by very small numbers.

Region-based analysis was also performed to assess the agreement of the tracer uptake and 28 radiomic features between predicted and ground-truth images. Using the PMOD medical image analysis software (PMOD Technologies LLC, Switzerland) and the Hammers N30R83 brain atlas, 83 brain regions were delineated on the ground-truth FD PET images. Then the delineated volumes regions were mapped to LD, PIS and PSS PET images to quantify 28 radiometric features using the LIFEx analysis tool (*35*). Moreover, the region-wise standardized uptake value (SUV) bias and standard deviation (STD) were calculated for the 83 brain regions on the predicted as well as LD PET images with the FD PET images serving as reference. A joint histogram analysis was also carried out to depict the voxel-wise correlation of the activity concentration between PIS and PSS and reference FD PET images.

Overall, 28 radiomic features were extracted for each brain regions including seven conventional indices, five first-order features, seven Grey-Level Zone Length Matrix (GLZLM), six Grey-Level Run Length Matrix (GLRLM) and three Grey Level Co-occurrence Matrix (GLCM) features. The list of these radiomic features is shown in Table 2. The relative error (RE%) was also calculated for the radiomic features using Eq. 6.

$$RE = \frac{(PIS,PSS)_f - FD_f}{FD_f} \times 100\% \quad (6)$$

In Eq. (6) , f denotes the value of a specific radiomic feature calculated in a brain region. The MedCalc software (*36*) was employed for the calculation of the pairwise t-test for statistical analysis of RMSE, SSIM and PSNR between LD, PSS, PIS and reference FD PET images. The significance level was set at *p*-value < 0.05 for all comparisons.

RESULTS

The predicted images in both image and projection space exhibited notable enhancement in image quality compared to LD images, providing almost similar appearance with respect to reference FD PET images. Figure 2 displays representative transverse, coronal and sagittal views showing reference FD, LD, PIS and PSS PET images along with their corresponding bias maps. The visual inspection revealed that the images derived from training in the sinogram space better reflected the underlying FDG uptake patterns and anatomy than those obtained from implementation in image space. The image quality scores assigned by

the two physicians to FD, LD, PIS, PSS images are shown in Figure 3. The mean scores for each group are indicated at the top of each bar. The PIS images were scored as poor (score = 2) or better. The FD and PSS images exhibited comparable quality with score of 4.9 and 4.55 (good) or higher, respectively.

Table 3 summarizes the PSNR, SSIM and RMSE calculated separately on the validation and test dataset for PIS, PSS and LD PET images. Overall, the predicted images in the projection-space showed improved image quality and better noise properties and higher quantitative accuracy (Table 4) with statistically significant differences with respect to the implementation in image-space.

Figure 4 illustrates linear regression plots depicting the correlation between tracer uptake for LD, PIS and PSS with respect to FD. The scatter and linear regression plots showed higher correlation between PSS and FD ($R^2 = 0.99$, RMSE = 0.28) compared to PIS ($R^2 = 0.98$, RMSE = 0.33). A relatively higher RMSE (0.42) was obtained for LD PET images.

The Bland-Altman plots, where each data point reflects a brain region, confirmed the results obtained from joint histogram analysis where the lowest SUV bias (0.02) and smallest SUV variance (95% CI: -0.92, +0.84) were observed for PSS images (figure 5). Though the SUV bias is extremely low for LD images, increased variance compared with FD images was observed (95% CI:-2, +2), reflecting their poor image quality.

Figure 6 compares the deep learning predicted images (PIS and PSS) and images reconstructed from the low dose sinogram using four state of the art iterative reconstruction methods, including OSEM, OSEM+TOF, OSEM+PSF and OSEM+TOF+PSF. The full-dose sinogram was also reconstructed using OSEM and OSEM+TOF+PSF as reference.

Supplemental Figure 2 depicts the region-wise quantitative accuracy of the tracer uptake for LD, PSS and PIS images. The standard deviation of tracer uptake in all brain regions (Supplemental Figure 3), SUV bias and its standard deviation within each brain region were calculated using Hammers' N30R83 brain atlas to delineate the 83 brain regions. It was shown that the SUV bias was below 4% for PSS, PIS and LD images with LD exhibiting a relatively high STD compared to PIS and PSS. The PSS approach led to the lowest absolute average SUV bias (0.69 ± 0.7%) across all brain regions, while PIS and LD resulted in absolute average SUV bias of 1.35 ± 1.15% and 1.12 ± 0.93%, respectively (Table 4). Even though a very low SUV bias was observed in LD images, a remarkably increased STD was seen, reflecting the high noise level in LD images. Symmetrical left and right sides of the brain regions were merged reporting a single value to reduce the number of regions. Hence, the 83 brain regions were reduced to 44 in figure 6. The higher standard deviation of SUV bias was observed in LD images, reflecting the noisy nature of low count images. Lower standard deviations were observed in PSS compared to PIS.

Supplemental Figures 4 and 5 shows the relative error (%) of 28 radiomic features calculated for PSS and PIS images across the 83 brain regions for the 20 subjects in the test dataset. The mean RE of SUV_{mean}

calculated across all brain regions was 0.24 ± 0.96% and 1.05 ± 1.44% for PSS and PIS, respectively. The largest SUV_{mean} bias between PSS and PIS images with respect to reference FD images were observed in the brainstem (4.04%), corp-callosum (3.8%), pallidum (3.08%), caudate nucleus (1.6%) and superior frontal gyrus (3.38%). SUV_{max} had a mean RE of 1.18 ± 1.5% and 0.81 ± 0.51% for PIS and PSS, respectively. The mean RE of the homogeneity radiomic feature belonging to GLCM category was -1.07 ± 1.77%, 0.28 ± 1.4% for PIS and PSS, respectively. Only 12 and 5 regions had a RE greater than 2% for PIS and PSS, respectively. The middle frontal gyrus, medial orbital gyrus, and posterior orbital gyrus displayed substantial variances for dissimilarity radiomic feature of both PIS and PSS (3.68% vs. 4.89%, -1.7% vs. 2.91% and -1.7% vs. 2.9%).

DISCUSSION

Table 5 summarizes the study design and outcomes of previous works reporting on the prediction of FD PET images from LD images based using deep learning approaches (*15-21*). In this work, we aimed to generate diagnostic quality FDG brain PET images from LD PET data in the image or projection domains corresponding to only just 5% of injected activity compared to the regular FD scan. The training of the neural network was performed using a so-called 2.5 D scheme, considering a batch of image slices as input, since there is a dependence of tracer distribution along the z-axis. Hence, by including the neighbouring slices, the model would be able to capture the underlying morphologic information. In contrast to previous studies, we aimed to train the network in projection and image domains to evaluate the performance of both approaches for estimation of FD PET images. It was shown that the synthesized FD images predicted from LD sinograms had a superior image quality and lower regional SUV bias and variance compared with both LD and FD images predicted in image-space. This highlights the value of employing raw data in sinogram space (400×168×621 =41'731'200) rather than the data in image space (101 × 101 × 71=724'271). The data representation in image-space is different from representation in projection-space. Let's consider an ideal point source located at the centre of the field-of-view, which would appear as a hot spot in the corresponding location in image space. The same point source would be reflected in sinogram space by numerous correlated lines of response, conveying different data representations of the same element. The extended/detailed data available in sinogram space helped the convolutional network to better decode the underlying features, thus resulting in superior performance. The convolutional network trained in the image domain applied simplistic noise reduction, thus leading to blurred, highly smoothed and slightly biased FD images.

The qualitative assessment of image quality performed by nuclear medicine specialists demonstrated the superior performance of the PSS approach, showing close agreement between PSS and the reference FD images. The RMSE calculated on LD and synthesized PSS and PIS images were 0.41 ± 0.03, 0.17 ± 0.01

and 0.18 ± 0.04, respectively, reflecting the effectiveness of model training in the sinogram space (p-value<0.05). Moreover, the SSIM improved from 0.84 ± 0.04 for LD to 0.96 ± 0.03 for PIS images and further to 0.97 ± 0.02 for PSS images. It would be enlightening to consider the resulting metrics in conjunction with those obtained from LD images for better interpretation of the extent of improvement achieved by the proposed methods. For instance, Ouyang *et al.* (*21*) claimed that only 1% of the standard dose was used yielding LD images with better or at least comparable SSIM (0.86 vs 0.84) and RMSE (0.2 vs 0.4) compared to ours with 5% of the full dose. This might partly stem from differences in terms of sensitivity between PET scanners, which directly affects the quality of the PET images. In this regards, previous studies conducted on the GE SIGNA PET/MRI (1, 2) took advantage of its higher sensitivity (21cps/kBq) and better count-rate performance characteristics (peak NEC of 210 kcps at 17.5 kBq/cc) compared to the Biograph mCT scanner used in this study with considerably lower detection sensitivity (9.7 cps/kBq) and count-rate performance (NEC = 180 kcps at 28 kBq/cc). Furthermore, their technique relied on support from coregistered MR images, which could partly explain why 5% LD image in the present study and 1% LD in abovementioned works resulted in comparable RMSE (~ 0.15).

The quantitative analysis of 83 brain regions in terms of 28 radiomic features showed high repeatability of the radiomic features for both PSS and PIS techniques. From the 2324 data points corresponding the to the number of regions multiplied by the number of radiomic features, only 3 and 9 data points for PIS and PSS, respectively, had a RE larger than 5%, with the remaining data points exhibiting no significant REs. The quantitative evaluation showed less than 1% mean absolute error in most brain regions for PSS. We involved both normal and abnormal patients to offer a heterogeneous dataset. Neurologic abnormalities present in our dataset included patients presenting with cognitive symptoms of possible neurodegenerative disease. Since the dataset for the training contained both normal and abnormal patients, data augmentation was applied to avoid overfitting and guarantee robust and effective training. The Bland & Altman analysis showed lower bias and variance in the 83 regional SUV_{mean} values obtained from PSS and PIS PET images compared with LD images. The Bland & Altman plots further demonstrated the superior performance of the PSS approach, resulting in SUV values that are comparable to the original FD images.

In terms of computation time, the training in the image domain is less demanding than training in the projection-space. Training in the image domain took ~38h versus ~210h in the sinogram space. Moreover, synthesis of a single PET image (after training) in the image domain takes ~100 seconds versus ~370 seconds required in the sinogram space. This stems from the increased data size and consequently added processing burden for the sinogram space implementation.

One of the limitations of the present study was that during the clinical evaluation, the LD images were relatively easy to be identified by physicians. Hence, they could have been subconsciously biased and assigned lower scores to these images. Moreover, patient motion during the PET/CT scan, particularly for

patients suffering from dementia that are more susceptible to involuntary motion, may impair image quality of both LD and FD PET images. However, motion might affect LD and FD PET images differently since the randomly selected events for creation of the LD images might not exactly follow the same motion pattern in the FD PET data.

CONCLUSION

We have demonstrated that high quality FDG brain PET images can be generated using deep learning approaches either in the image domain or projection (sinogram) space. The noise was effectively reduced in the predicted PET images from the LD data. Prediction of FD PET images in the sinogram space exhibited superior performance, resulting in higher image quality and minimal quantification bias.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation under grants SNRF 320030_176052 and SNF 320030_169876 and the Swiss Cancer Research Foundation under Grant KFS-3855-02-2016.

Conflict of interest statement

The authors declare that they have no conflict of interest.

KEY POINTS

QUESTION: Does implementation of deep learning-guided low-dose brain ^{18}F-FDG PET imaging in projection-space improves performance over implementation in image-space?

PERTINENT FINDINGS: Using a cohort study comparing 140 clinical brain ^{18}F-FDG PET/CT studies, among which 100, 20 and 20 patients were randomly partitioned into training, validation and independent validation sets, we demonstrate through qualitative assessment and quantitative analysis demonstrated that the FD PET prediction in projection space led to superior performance, resulting in higher image quality and lower SUV bias and variance compared to FD PET prediction in the image domain.

IMPLICATIONS FOR PATIENT CARE: The proposed deep learning-guided denoising technique enables substantial reduction of radiation dose to patients and is applicable in a clinical setting.

REFERENCES

1. Zaidi H, Montandon M-L, Assal F. Structure-function based quantitative brain image analysis. *PET Clinics.* 2010;5:155-168.

2. NAS/NRC (National Academy of Sciences/National Research Council). *Health risks from exposure to low levels of ionizing radiation: BEIR VII – Phase 2.* Washington DC: National Research Council; 2006: pp 155-189.

3. Gatidis S, Wurslin C, Seith F, et al. Towards tracer dose reduction in PET studies: Simulation of dose reduction by retrospective randomized undersampling of list-mode data. *Hell J Nucl Med.* 2016;19:15-18.

4. Reader AJ, Zaidi H. Advances in PET image reconstruction. *PET Clinics.* 2007;2:173-190.

5. Yu S, Muhammed HH. Comparison of pre-and post-reconstruction denoising approaches in positron emission tomography. Paper presented at the 1st International Conference on Biomedical Engineering (IBIOMED), 5-6 October 2016; Yogyakarta, Indonesia, pp 1-6. doi: 10.1109/IBIOMED.2016.786982

6. Arabi H, Zaidi H. Improvement of image quality in PET using post-reconstruction hybrid spatial-frequency domain filtering. *Phys Med Biol.* 2018;63:215010.

7. Chenye W, Zhenghui H, Pengcheng S, Huafeng L. Low dose PET reconstruction with total variation regularization. *Conf Proc IEEE Eng Med Biol Soc.* 2014;2014:1917-1920.

8. Wang Y, Zhang P, An L, et al. Predicting standard-dose PET image from low-dose PET and multimodal MR images using mapping-based sparse representation. *Phys Med Biol.* 2016;61:791-812.

9. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.

10. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health.* 2019 *in press*

11. Arabi H, Zeng G, Zheng G, Zaidi H. Novel adversarial semantic structure deep learning for MRI-guided attenuation correction in brain PET/MRI. *Eur J Nucl Med Mol Imaging.* 2019 *in press*

12. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys.* 2017;44:1408-1419.

13. Bortolin K, Arabi H, Zaidi H. Deep learning-guided attenuation and scatter correction in brain PET/MRI without using anatomical images. Paper presented at: IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC); 26 October - 2 November 2019, 2019; Manchester, UK.

14. Yang J, Park D, Gullberg GT, Seo Y. Joint correction of attenuation and scatter in image space using deep convolutional neural networks for dedicated brain (18)F-FDG PET. *Phys Med Biol.* 2019;64:075019.

15. Chen KT, Gong E, de Carvalho Macruz FB, et al. Ultra-low-dose (18)F-Florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. *Radiology.* 2019;290:649-656.

16. Kang J, Gao Y, Shi F, Lalush DS, Lin W, Shen D. Prediction of standard-dose brain PET image by using MRI and low-dose brain [18F]FDG PET images. *Med Phys.* 2015;42:5301-5309.

17. Kaplan S, Zhu YM. Full-dose PET image estimation from low-dose PET image using deep learning: a pilot study. *J Digit Imaging.* 2019;32:773-778.

18. Schaefferkoetter JD, Yan J, Soderlund TA, et al. Quantitative accuracy and lesion detectability of low-dose FDG-PET for lung cancer screening. *J Nucl Med.* 2017;58:399-405.

19. Xiang L, Qiao Y, Nie D, An L, Wang Q, Shen D. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. *Neurocomputing.* 2017;267:406-416.

20. Wang Y, Yu B, Wang L, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *Neuroimage.* 2018;174:550-562.

21. Ouyang J, Chen KT, Gong E, Pauly J, Zaharchuk G. Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. *Med Phys.* 2019;46:3555-3564.

22. Xu J, Gong E, Pauly J, Zaharchuk G. 200x Low-dose PET reconstruction using deep learning. *ARXIV.* 2017:eprint arXiv:1712.04119.

23. Haggstrom I, Schmidtlein CR, Campanella G, Fuchs TJ. DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem. *Med Image Anal.* 2019;54:253-262.

24. Hong X, Zan Y, Weng F, Tao W, Peng Q, Huang Q. Enhancing the image quality via transferred deep residual learning of coarse PET sinograms. *IEEE Trans Med Imaging.* 2018;37:2322-2332.

25. Liu CC, Qi J. Higher SNR PET image prediction using a deep learning model and MRI image. *Phys Med Biol.* 2019;64:115004.

26. Cui J, Gong K, Guo N, et al. PET image denoising using unsupervised deep learning. *Eur J Nucl Med Mol Imaging.* 2019: *in press*

27. Lu W, Onofrey JA, Lu Y, et al. An investigation of quantitative accuracy for deep learning based denoising in oncological PET. *Phys Med Biol.* 2019;64:165019.

28. Schaefferkoetter J, Nai YH, Reilhac A, Townsend DW, Eriksson L, Conti M. Low dose positron emission tomography emulation from decimated high statistics: A clinical validation study. *Med Phys.* 2019;46:2638-2645.

29. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical image computing and computer-assisted intervention, 2015; 9351:234-241.

30. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. International conference on medical image computing and computer-assisted intervention, 2016; pp 424-432.

31. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; pp 4510-4520.

32. Chollet F. Keras: The python deep learning library. *Astrophysics Source Code Library.* 2018.

33. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980.* 2014.

34. Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association.* 1987;82:528-540.

35. Nioche C, Orlhac F, Boughdad S, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer research.* 2018;78:4786-4789.

36. Schoonjans F, Zalata A, Depuydt C, Comhaire F. MedCalc: a new computer program for medical statistics. *Computer methods and programs in biomedicine.* 1995;48:257-262.

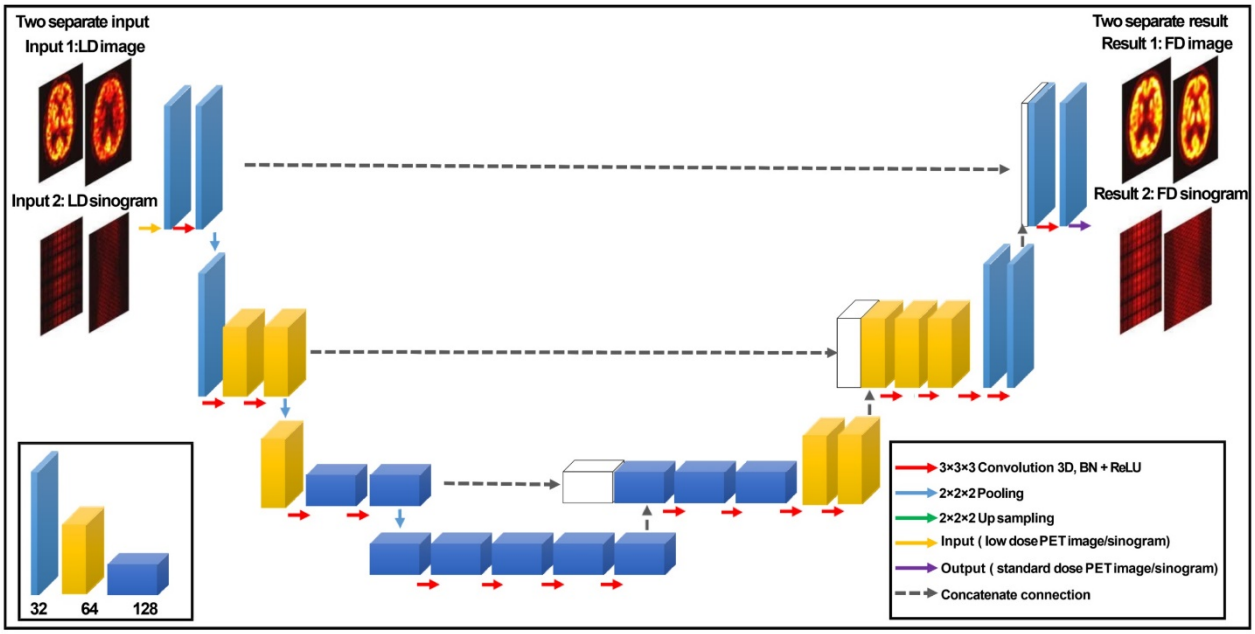


FIGURE 1. Schematic diagram of the modified 3D U-net, consisting of an encoder-decoder convolutional neural network. The tensors are indicated by the boxes whereas the arrows denote the computational operations. BN = batch normalization, ReLU = rectified linear unit activation. The number of channels is indicated beneath each box in the bottom left panel. The input and output of this network are LD and FD image PET pairs either in image or sinogram space.

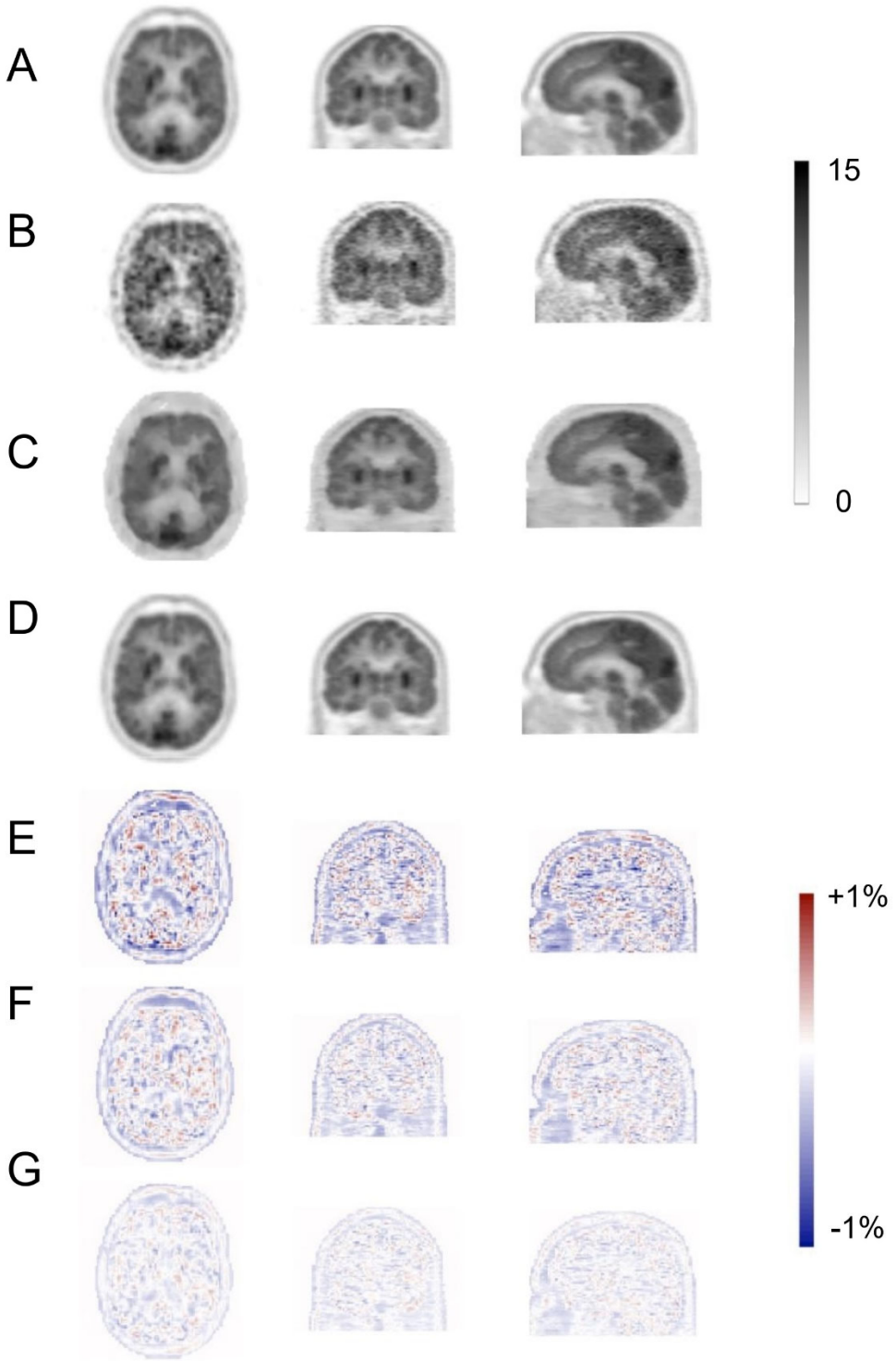


FIGURE 2. Representative FDG brain PET image of a 65-year old male patient. The reference full-dose and the corresponding low-dose and predicted FD images in the image and sinogram space are presented. SUV bias maps for LD, PIS and PSS PET images with respect to the reference FD PET image are shown in the lower panel.

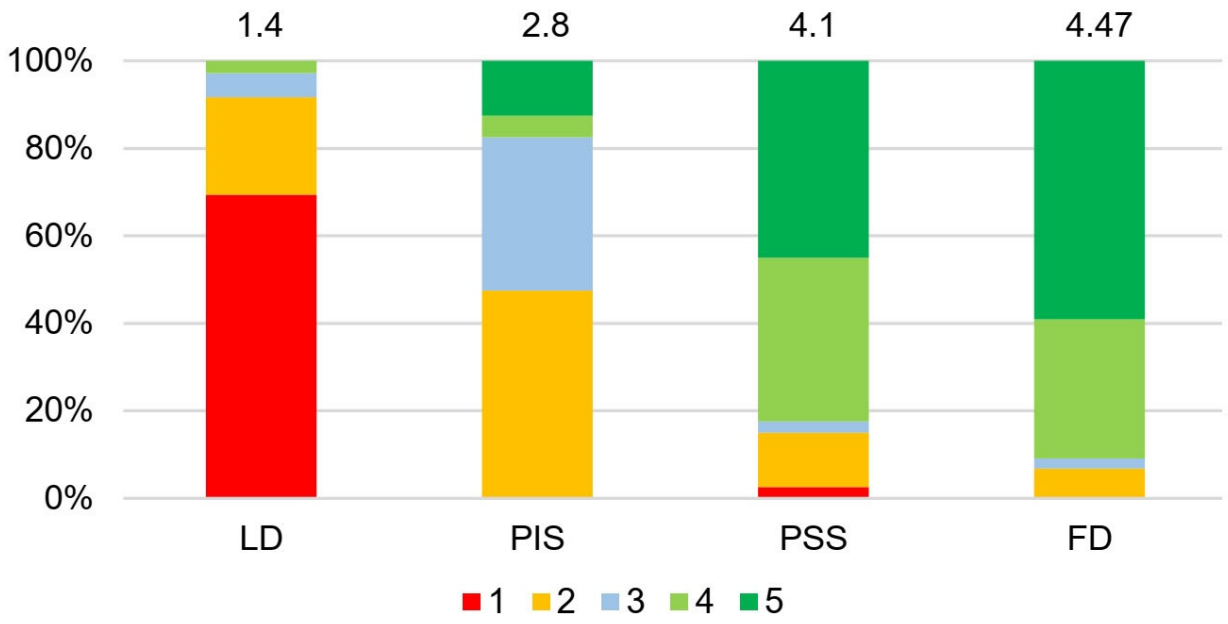


FIGURE 3. Result of image quality assessment by the two nuclear medicine specialists for LD, PIS, PSS and FD PET images. Mean scores are presented on the top of the bar plots. 1 = uninterpretable, 2 = poor, 3 = adequate, 4 = good, 5 = excellent.

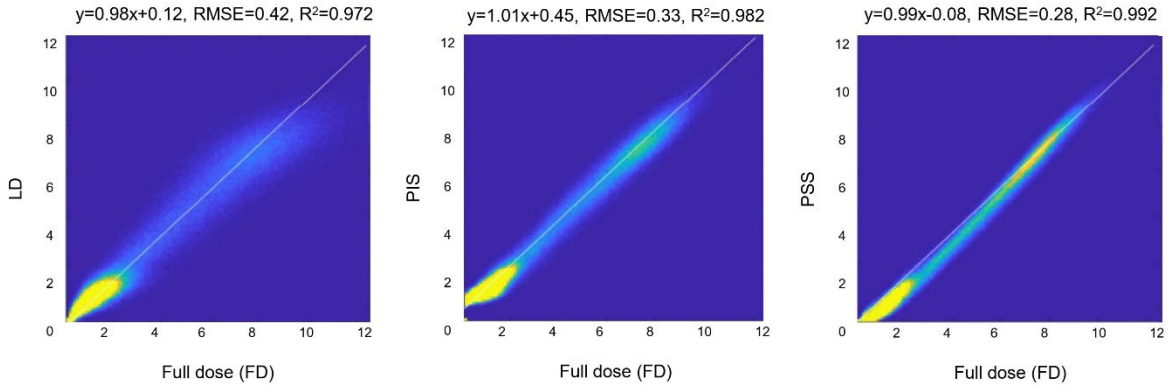


FIGURE 4. Joint histogram analysis of the LD PET images (left), predicted FD images in sinogram space (middle) and image space (right) versus FD PET images.

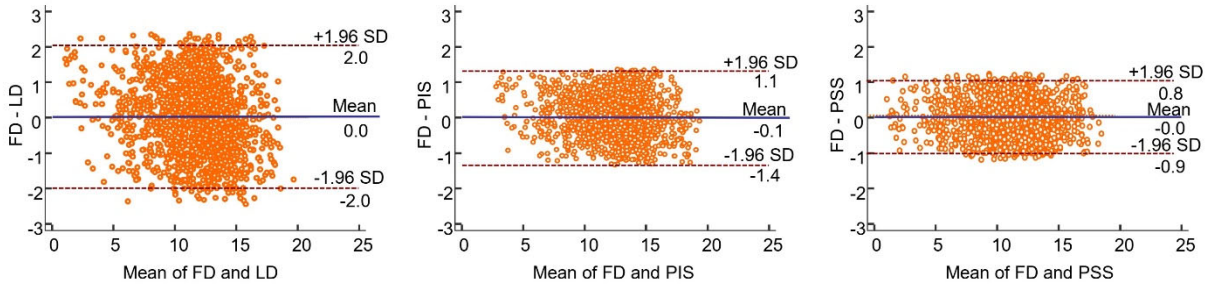


FIGURE 5. Bland & Altman plots of SUV differences in the 83 brain regions calculated for LD (left), PIS (middle) and PSS (right) PET images with respect to the reference FD PET images in the test dataset. The solid blue and dashed lines denote the mean and 95% confidence interval (CI) of the SUV differences, respectively.

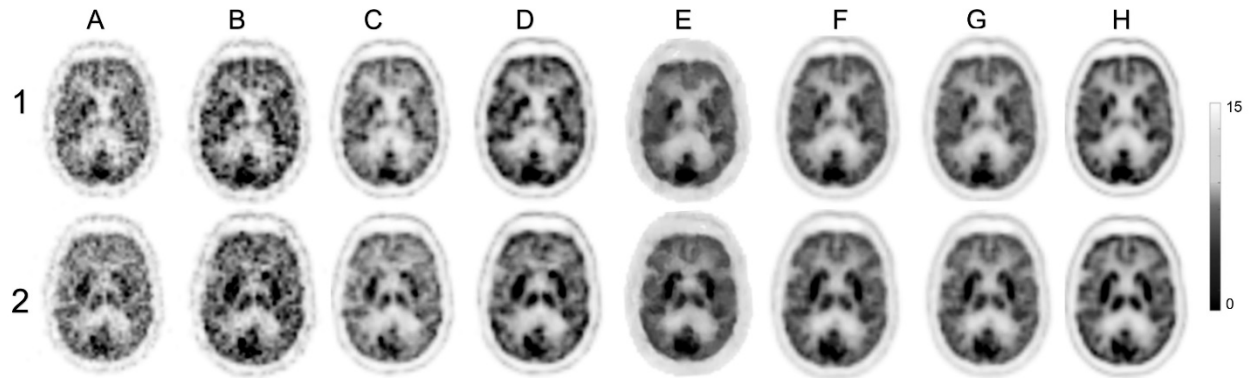


FIGURE 6. Comparison of images of two clinical ^{18}F-FDG brain PET studies (1 and 2) reconstructed from 5% low dose sinograms using four different reconstruction algorithms, including (A) OSEM, (B) OSEM+TOF, (C) OSEM+PSF and (D) OSEM+TOF+PSF with deep learning-based predicted images in image space (E) and sinogram space (F). Reference full dose images reconstructed using (G) OSEM and (H) OSEM+TOF+PSF are also shown.

Table 1. Demographics of patients included in this study.

	Training	Test	Validation
Number	100	20	20
Male/Female	45/55	11/9	8/12
Age (Mean ± SD)	73±8	68±18	73±4.5
Weight(Mean ± SD)	70±13	67±12	71±11
Indication/Diagnosis	Cognitive symptoms of possible neurodegenerative etiology		

Table 2. Summary of the 28 radiomic features belonging to the 6 main categories estimated for the 83 brain regions.

Radiomic feature category	Radiomic feature
Conventional indices	SUV_{mean} SUV_{std} SUV_{max} SUV Q1 SUV Q2 SUV Q3 TLG (mL)
First-order features - Histogram	Kurtosis Entropy_log10 Entropy_log2
First-order features - Shape	SHAPE_Volume (ml) SHAPE_Volume (# Voxel)
Grey-Level Zone Length Matrix (GLZLM)	Short-Zone Emphasis (SZE) Long-Zone Emphasis (LZE) Short-Zone Low Gray-level Emphasis (SZLGE) Short-Zone High Gray-level Emphasis (SZHGLE) Long-Zone Low Gray-level Emphasis (LZLGLE) Long-Zone High Gray-level Emphasis (LZHGLE) Zone Percentage (ZP)
Grey-Level Run Length Matrix (GLRLM)	Short-Run Emphasis (SRE) Long-Run Emphasis (LRE) Short-Run Low Gray-level Emphasis (SRLGLE) Short-Run High Gray-level Emphasis(SRHGLE) Run Length Non-Uniformity (RLNU) Run Percentage (RP)
Grey Level Co-occurrence Matrix (GLCM)	Homogeneity Energy Dissimilarity

Table 3. Comparison of the results obtained from analysis of image quality in LD PET images and predicted images in image (PIS) and sinogram (PSS) space for the validation dataset. SSIM: structural similarity index metrics, PSNR: peak signal to noise ratio, RMSE: root mean squared error.

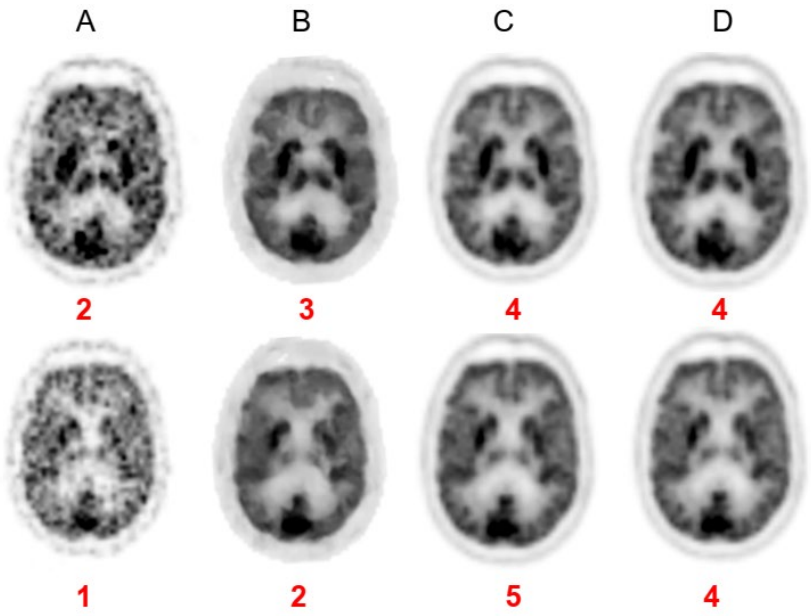
Validation dataset	SSIM	PSNR	RMSE
PIS	0.97±0.02	34.60±1.08	0.18±0.02
PSS	0.98±0.01	38.25±0.66	0.15±0.03
LD	0.84±0.04	29.00±0.92	0.40±0.03
P-value (PIS vs. PSS)	0.022	0.019	0.016
P-value (PIS vs. LD)	0.037	0.021	0.036
P-value (PSS vs. LD)	0.042	0.025	0.030
Test dataset	**SSIM**	**PSNR**	**RMSE**
PIS	0.96±0.03	31.70±0.75	0.18±0.04
PSS	0.97±0.02	37.30±0.71	0.17±0.01
LD	0.82±0.15	29.92±0.71	0.41±0.04
P-value (PIS vs. PSS)	0.031	0.024	0.018
P-value (PIS vs. LD)	0.040	0.036	0.041
P-value (PSS vs. LD)	0.041	0.031	0.031

Table 4. Average and absolute average SUV bias ± STD calculated across the 83 brain regions for LD, PIS, and PSS PET images.

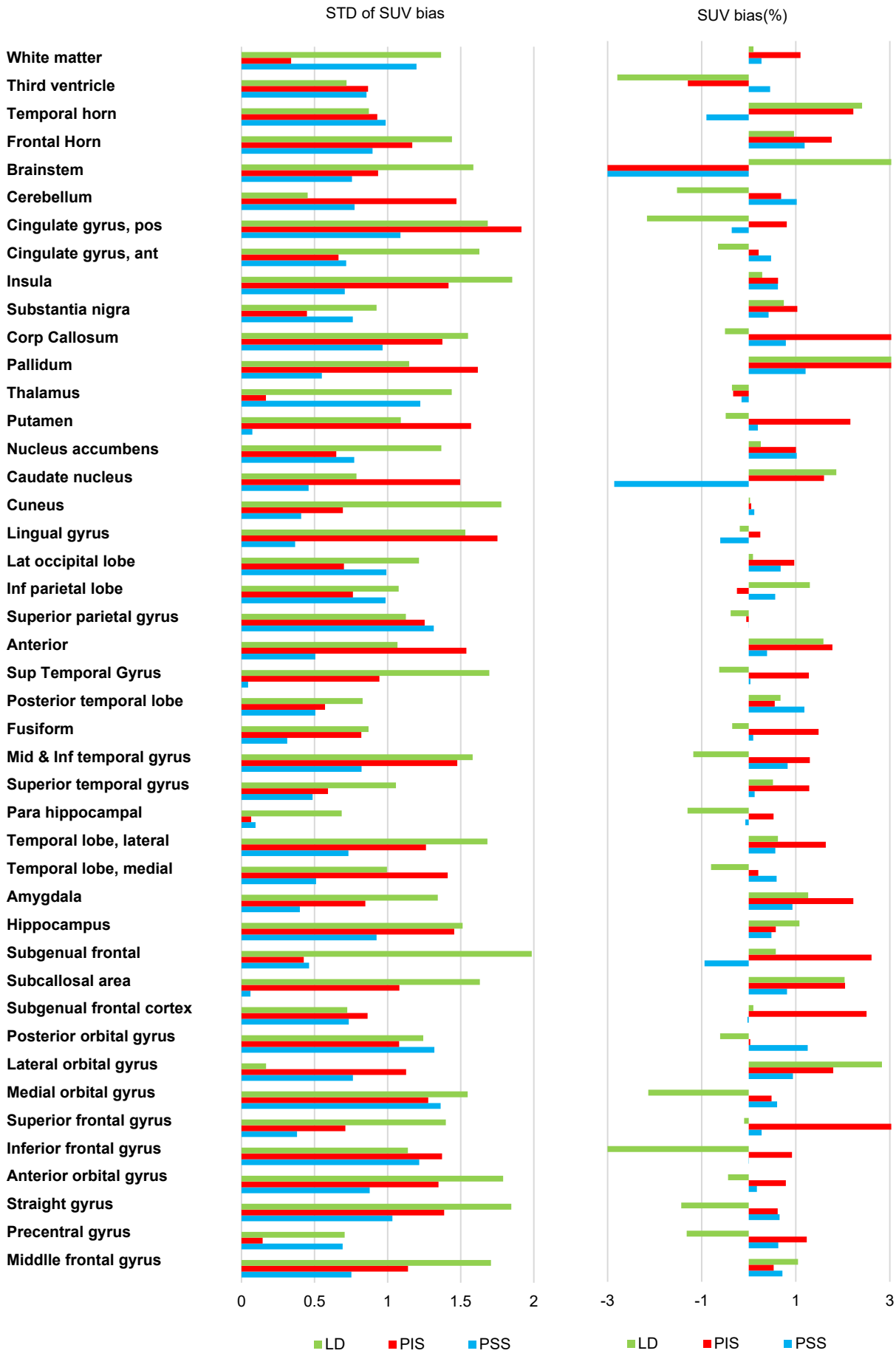
	PSS	PIS	LD
Average SUV bias (%)	0.24±0.96	1.05±1.44	0.10±1.47
Absolute average SUV bias (%)	0.69±0.70	1.35±1.15	1.12±0.93

Table 5. Summary of previous studies focusing on LD to FD PET conversion in brain PET imaging using deep learning techniques.

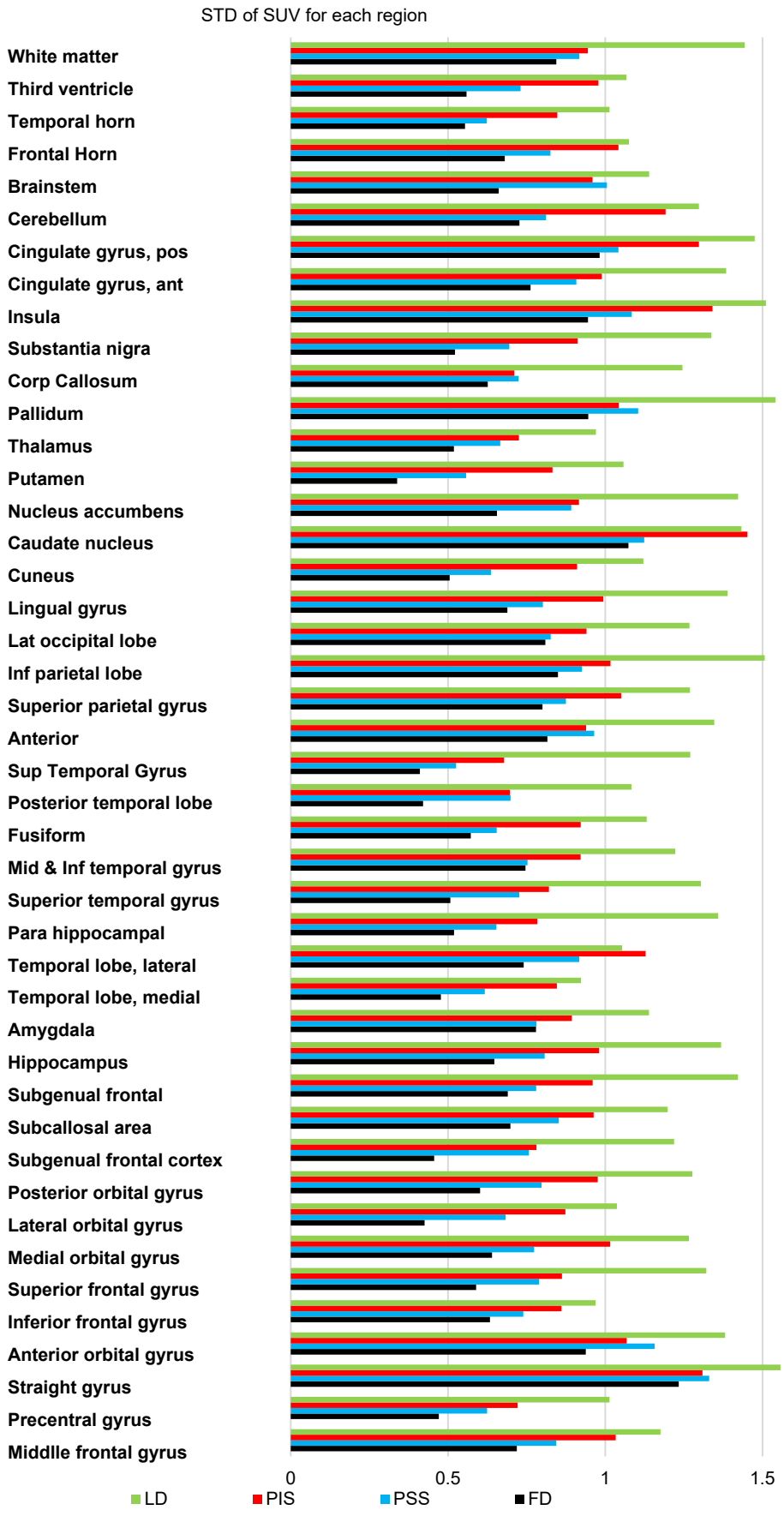
Reference	Nb. patients	Input	Network architecture	Injected activity (MBq)	Low-dose ratio (%)	Tracer	Scanning time (min)	Time from injection to scan (min)	Scanner model	Evaluation method and metrics
(*16*)	11	LD PET, MRI (T1)	Segmented brain tissues from MRI to build tissue-specific models to predict standard-dose plus an iterative refinement strategy	203±12	25	^{18}F-FDG	12	36	Siemens Biograph mMR	Leave-one-out cross-validation, SUV bias
(*8*)	8	LD PET, MRI (T1, DTI)	Sparse representation based on mapping strategy and an incremental refinement scheme	203±12	25	^{18}F-FDG	12	60	Siemens Biograph mMR	NMSE,PSNR, contrast recovery, quantification bias
(*19*)	16	LD PET MRI (T1)	Deep auto-context CNN architecture with three 4-layers	203±12	25	^{18}F-FDG	12	60	Siemens Biograph mMR	NMSE, PSNR, Training and Testing time comparison, Training Loss vs. Iteration
(*22*)	9	LD PET	2.5D U-Net (modified)	370	0.5	^{18}F-FDG	40	45	GE SIGNA PET/MRI	NRMSE, error maps
(*15*)	40	LD PET MRI (T1, T2, DTI)	2D U-Net	330±30	1	^{18}F-Florbetaben	20	90-110	GE SIGNA PET/MRI	SSIM, PSNR, RMSE, clinical image quality scoring, Bland-Altman analysis
(*20*)	16	LD PET	3D conditional GANs	203±12	25	^{18}F-FDG	12	36	Siemens Biograph mMR	SSIM, PSNR, RMSE, quantification bias
(*21*)	40	LD PET	2.5D GAN	330±30	1	^{18}F-Florbetaben	20	90-111	GE SIGNA PET/MRI	SSIM, PSNR, RMSE, clinical image quality scoring, error maps
This work in image space	140	LD PET image	2.5D U-Net (modified)	205±10	5	^{18}F-FDG	20	34	Siemens Biograph mCT	SSIM, PSNR, RMSE, clinical image quality scoring, Bland-Altman analysis, quantification bias, radiomic features, joint histogram, error maps
This work in projection space	140	LD PET sinograms	2.5D U-Net (modified)	205±10	5	^{18}F-FDG	20	34	Siemens Biograph mCT	SSIM, PSNR, RMSE, clinical image quality scoring, Bland-Altman analysis, quantification bias, radiomic features, joint histogram, error maps



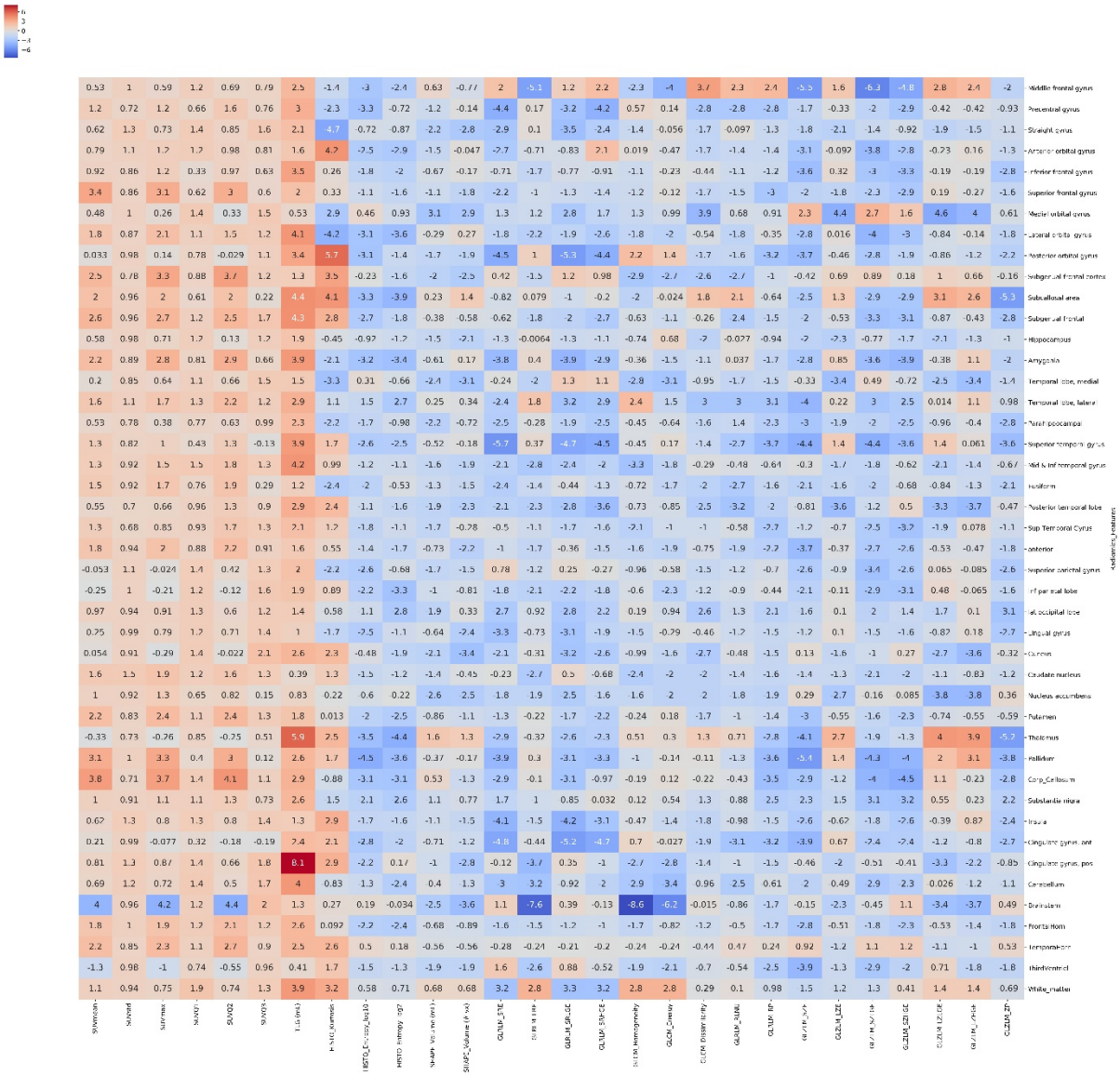
Supplemental Figure 1. Representative clinical ^{18}F-FDG brain PET images from two extremes of image evaluation. The images shown are (A) low-dose, (B) deep learning prediction in image space, (C) deep learning prediction in sinogram space, (D) full-dose. The red numbers indicate scores assigned by one expert physician to images.


STD of SUV bias
SUV bias(%)
White matter
Third ventricle
Temporal horn
Frontal Horn
Brainstem
Cerebellum
Cingulate gyrus, pos
Cingulate gyrus, ant
Insula
Substantia nigra
Corp Callosum
Pallidum
Thalamus
Putamen
Nucleus accumbens
Caudate nucleus
Cuneus
Lingual gyrus
Lat occipital lobe
Inf parietal lobe
Superior parietal gyrus
Anterior
Sup Temporal Gyrus
Posterior temporal lobe
Fusiform
Mid & Inf temporal gyrus
Superior temporal gyrus
Para hippocampal
Temporal lobe, lateral
Temporal lobe, medial
Amygdala
Hippocampus
Subgenual frontal
Subcallosal area
Subgenual frontal cortex
Posterior orbital gyrus
Lateral orbital gyrus
Medial orbital gyrus
Superior frontal gyrus
Inferior frontal gyrus
Anterior orbital gyrus
Straight gyrus
Precentral gyrus
Middlle frontal gyrus
0
0.5
1
1.5
2
-3
-1
1
3
LD
PIS
PSS
LD
PIS
PSS

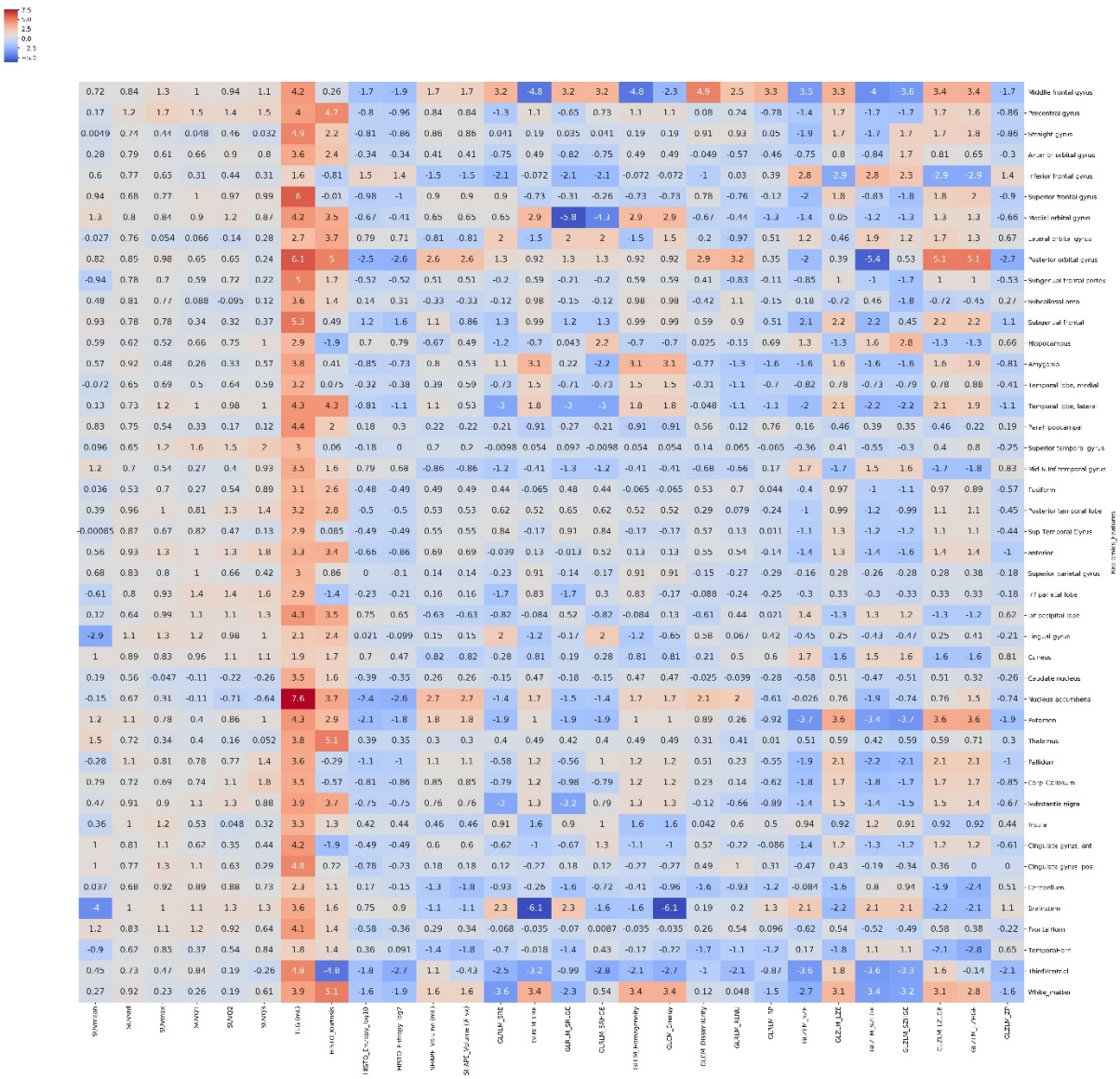
Supplemental Figure 2. Plots of SUV bias standard deviation (left panel) and mean SUV bias (%) (right panel) in the different brain regions for LD, PIS, and PSS PET images. The left and right regions were merged, thus reducing the number of brain regions to 44.


STD of SUV for each region
White matter
Third ventricle
Temporal horn
Frontal Horn
Brainstem
Cerebellum
Cingulate gyrus, pos
Cingulate gyrus, ant
Insula
Substantia nigra
Corp Callosum
Pallidum
Thalamus
Putamen
Nucleus accumbens
Caudate nucleus
Cuneus
Lingual gyrus
Lat occipital lobe
Inf parietal lobe
Superior parietal gyrus
Anterior
Sup Temporal Gyrus
Posterior temporal lobe
Fusiform
Mid & Inf temporal gyrus
Superior temporal gyrus
Para hippocampal
Temporal lobe, lateral
Temporal lobe, medial
Amygdala
Hippocampus
Subgenual frontal
Subcallosal area
Subgenual frontal cortex
Posterior orbital gyrus
Lateral orbital gyrus
Medial orbital gyrus
Superior frontal gyrus
Inferior frontal gyrus
Anterior orbital gyrus
Straight gyrus
Precentral gyrus
Middlle frontal gyrus
0
0.5
1
1.5
LD
PIS
PSS
FD

Supplemental Figure 3. Plot of SUV standard deviation in the different brain regions for LD, PIS, and PSS PET images. The left and right regions were merged, thus reducing the number of brain regions to 44.



Supplemental Figure 4. Heat map of the relative error of 28 radiomics features calculated across 83 brain regions for PIS PET images with respect to reference FD PET images.



Supplemental Figure 5. Heat map of the relative error of 28 radiomics features calculated across 83 brain regions for PSS PET images with respect to reference FD PET images.