EJNMMI Physics

ORIGINAL RESEARCH

Open Access

Transfer learning-based attenuation correction in ^{99m}Tc-TRODAT-1 SPECT for Parkinson's disease using realistic simulation and clinical data

Wenbo Huang[1], Han Jiang[2], Yu Du[1,3], Haiyan Wang[1,4], Hao Sun[5], Guang-Uei Hung[6] and Greta S. P. Mok[1,3*]

*Correspondence:
Greta S. P. Mok
gretamok@um.edu.mo
[1]Biomedical Imaging Laboratory (BIG), Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Macau, Taipa, Macau SAR, China
[2]PET-CT Center, Fujian Medical University Union Hospital, Fuzhou, China
[3]Center for Cognitive and Brain Sciences, Institute of Collaborative Innovation, University of Macau, Taipa, Macau SAR, China
[4]Research Center for Medical AI, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[5]School of Biomedical Engineering, Southern Medical University, Guangzhou, China
[6]Department of Nuclear Medicine, Chang Bing Show Chwan Memorial Hospital, Changhua County, Taiwan

Abstract

Purpose Dopamine transporter (DAT) SPECT is an effective tool for early Parkinson's disease (PD) detection and heavily hampered by attenuation. Attenuation correction (AC) is the most important correction among other corrections. Transfer learning (TL) with fine-tuning (FT) a pre-trained model has shown potential in enhancing deep learning (DL)-based AC methods. In this study, we investigate leveraging realistic Monte Carlo (MC) simulation data to create a pre-trained model for TL-based AC (TLAC) to improve AC performance for DAT SPECT.

Methods A total number of 200 digital brain phantoms with realistic ^{99m}Tc-TRODAT-1 distribution was used to generate realistic noisy SPECT projections using MC SIMIND program and an analytical projector. One hundred real clinical ^{99m}Tc-TRODAT-1 brain SPECT data were also retrospectively analyzed. All projections were reconstructed with and without CT-based attenuation correction (CTAC/NAC). A 3D conditional generative adversarial network (cGAN) was pre-trained using 200 pairs of simulated NAC and CTAC SPECT data. Subsequently, 8, 24, and 80 pairs of clinical NAC and CTAC DAT SPECT data were employed to fine-tune the pre-trained U-Net generator of cGAN (TLAC-MC). Comparisons were made against without FT (DLAC-MC), training on purely limited clinical data (DLAC-CLI), clinical data with data augmentation (DLAC-AUG), mixed MC and clinical data (DLAC-MIX), TL using analytical simulation data (TLAC-ANA), and Chang's AC (ChangAC). All datasets used for DL-based methods were split to 7/8 for training and 1/8 for validation, and a 1-/2-/5-fold cross-validation were applied to test all 100 clinical datasets, depending on the numbers of clinical data used in the training model.

Results With 8 available clinical datasets, TLAC-MC achieved the best result in Normalized Mean Squared Error (NMSE) and Structural Similarity Index Measure (SSIM) (TLAC-MC; NMSE = 0.0143 ± 0.0082/SSIM = 0.9355 ± 0.0203), followed by DLAC-AUG, DLAC-MIX, TLAC-ANA, DLAC-CLI, DLAC-MC, ChangAC and NAC. Similar trends exist when increasing the number of clinical datasets. For TL-based AC methods, the fewer clinical datasets available for FT, the greater the improvement as compared to DLAC-CLI using the same number of clinical datasets for training. Joint histograms analysis and Bland-Altman plots of SBR results also demonstrate consistent findings.

Springer Open

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

Conclusion TLAC is feasible for DAT SPECT with a pre-trained model generated purely based on simulation data. TLAC-MC demonstrates superior performance over other DL-based AC methods, particularly when limited clinical datasets are available. The closer the pre-training data is to the target domain, the better the performance of the TLAC model.

Keywords Parkinson's disease, Brain SPECT, Transfer learning, Monte carlo simulation, Attenuation correction, ^{99m}Tc-TRODAT-1

Introduction

Parkinson's Disease (PD) is the second most prevalent and the fastest increasing neurodegenerative disorder, imposing an increasingly heavy burden on both families and societies [1]. PD is characterized primarily by the progressive degeneration of dopaminergic neurons containing neuromelanin in the substantia nigra, resulting in a significant reduction in dopamine transporter (DAT) levels within the nigrostriatal pathway [2]. Currently, single photon emission computed tomography (SPECT), is a well-established non-invasive imaging method for PD diagnosis, with ^{123}I-ioflupane [3] and ^{99m}Tc-TRODAT-1 [4] as two commonly used radionuclide tracers to target DAT.

However, several image degradation factors, including attenuation, scatter and collimator septal penetration [5], affect the image quality and quantitative accuracy in SPECT. Among these factors, attenuation has the most substantial impact [6], making attenuation correction (AC) an essential process to enhance SPECT image quality and quantitative accuracy. In clinical practice, CT-based attenuation correction (CTAC) has been demonstrated to compensate for the attenuation effect in DAT SPECT effectively, improving both image quality and quantitative accuracy [7]. In hybrid SPECT-CT systems, attenuation maps (μ-maps) can be derived from CT scans and incorporated into the iterative reconstruction to generate AC images [8]. However, stand-alone SPECT systems currently occupy a majority (~80%) of the SPECT market share [9]. Even when CT is available, the presence of CT artifacts [10] and misregistration between CT and SPECT [11] can lead to attenuation artifacts. Additionally, the radiation exposure from CT may also increase the risk of cancer [12]. Chang's AC method [13] (ChangAC) is a conventional CT-less AC method that assumes homogenous tissue distribution. However, this assumption may lead to estimation errors, especially for areas with high μ values, such as the skull in brain SPECT.

Recently, deep learning (DL) has emerged as a promising alternative in SPECT AC [14]. Primary challenges associated with DL methods are the need for sufficient clinical data and the associated computation time for training. New training methods with relatively limited clinical data are of clinical interest to enhance the practicability. Transfer learning (TL) emerges as a promising DL strategy, involving fine-tuning (FT) a pre-trained model with reduced data and training time [15]. For nuclear medicine imaging, Sun et al. [16] demonstrated the application of TL for cross-tracer and cross-scanner AC in brain SPECT. However, TL strategies still need extensive clinical data for pre-training, which could be difficult to obtain as well.

Digital phantoms [17], coupled with Monte Carlo (MC) simulations [18], can generate realistic data and serve as evaluation benchmarks without physical experimentation or prototype manufacturing [19–22]. The introduction of a realistic phantom population [23] further enhances the feasibility of a virtual clinical trial [24]. Recently, we proposed

a method to generate a realistic digital brain phantom population for PD that includes variations in anatomical features and DAT tracer uptake according to the SPECT Visual Interpretation (VI) assessment scheme [25]. MC simulated data have been applied on DL-based image processing in SPECT [26], yet no study has explored the feasibility of a simulation-based pre-trained model for brain SPECT TLAC. In this work, we develop a 3D conditional generative adversarial network (cGAN) to estimate the CTAC images from non-attenuation-corrected (NAC) images of DAT SPECT, using the TL strategy to fine-tune a simulation-based pre-trained model with limited patient data. We also evaluate different DL and TL training strategies and the use of different number of FT data on the model performance.

Materials and methods

Data acquisition and preprocessing

Clinical dataset

^{99m}Tc-TRODAT-1 SPECT data of 100 patients with suspected PD (age range: 35 to 87 years, mean ± standard deviation (SD): 70.88 ± 9.84 years; injected activity: 1110 MBq; 55 male and 45 female) were retrospectively collected in 2016 from Show Chwan Memorial Hospital under the local ethics approval (SCMH_IRB Number, 1311110704). All data were acquired using a commercial clinical dual-head SPECT/CT scanner (Symbia, Siemens Healthineers, Erlangen, Germany) equipped with 2 low-energy high-resolution parallel-hole collimators, and a 9.6 mm (3/8 inch) thick NaI crystal with a 3.8 mm intrinsic resolution. The dual energy window (DEW) scatter correction method was implemented, with the primary and scatter windows set at 126–154 keV and 114–126 keV, respectively. The SPECT data comprised of 120 projection views with a bin size of 2.7 mm. CT images were acquired for SPECT AC after the SPECT scan using a 2-slice CT scanner with acquisition parameters of 10 mAs, 130 kVp, a pitch of 1.5, and a reconstructed slice thickness of 3 mm. The CT images were subsequently registered to SPECT data using the scanner software and then converted to μ-maps using a bilinear model [27].

Simulation dataset

We generated a population of 200 realistic digital PD phantoms based on our previous study [28]. The MC simulation tool SIMIND [19] was used to generate realistic simulated projection data modeling all relevant physical factors as well as the same SPECT system and data acquisition parameters as the clinical data. Almost noise-free projections with 120 views and a 1 mm bin size were obtained by about 50 M simulated photon counts. Projections were then resampled from $256 \times 200 \times 120$ to $95 \times 74 \times 120$ with a 2.7 mm bin size consistent with clinical data to model continuous-to-discrete activity sampling in clinical data acquisition [29]. We also generated a set of analytical simulation projections, modeling only the attenuation effect using a rotation based analytical projector [30]. We then scaled both MC and analytical simulation projections following a truncated Gaussian distribution according to our clinical reference (2.49 ± 0.73 M, range 0.49 ~ 4.88 M), and modeled Poisson noise on the scaled projections to obtain realistic noisy simulated projections.

The reconstruction process utilized ordered subset expectation maximization algorithm (8 iterations, 4 subsets). Both clinical and simulation projections were

reconstructed with and without CT-based AC (CTAC and NAC). Additionally, ChangAC for clinical data was implemented for comparison. DEW scatter correction and collimator-detector response correction were applied for both clinical and MC simulated data. The reconstructed images were with a matrix size of 128 × 128 × 128 and a voxel size of 2.7 × 2.7 × 2.7 mm^3. A Gaussian filter with a full-width-at-half-maximum of 4.8 mm was applied to all reconstructed images. Sample images of simulated and clinical CTAC data were presented in Supplementary Fig. 1.

Network architecture

In this study, a 3D cGAN was implemented [14], consisting of a U-Net generator (*G*) and a convolutional neural network (CNN) discriminator (*D*). The U-Net generator consisted of four encoder-decoder layers, each included two convolutional layers with a kernel size of 3 × 3 × 3, followed by a rectified linear unit (ReLU) activation function. After the activation function, a 3D max-pooling layer with a stride of 2 and a kernel size of 2 × 2 × 2 was used for each encoding step, and transposed convolutional layers were used to recover the feature maps for the decoding step. The recovered feature maps were concatenated with the skip connection from the corresponding encoder layer and served as input for the subsequent decoder layer. After the final decoder layer, a convolutional layer with a kernel size of 1 × 1 × 1 was employed to produce the final estimated data as the network output.

The discriminator consisted of four 3D convolutional layers with a stride of 2 and a kernel size of 3 × 3 × 3. Each convolutional layer was followed by a leaky ReLU (LReLU) activation function with a slope of 0.2. CTAC images and DL-based AC images from the generator were used as input. Then a binary classification was made to evaluate the authenticity of the input images. The generator loss L_G and the discriminator L_D were defined as follows:

$$L_G(x,y) = L_{adv}(x) + \lambda\, SL_1(G(x), y) \tag{1}$$

$$L_D(x,y) = \tfrac{1}{2}\left(\left(D(x,y) - T_{real}\right)^2 + \left(D(x, G(x) - T_{synthetic}\right)^2\right) \tag{2}$$

Where x and y were the NAC and CTAC SPECT image, respectively. L_{adv} was the adversarial loss function of the generator, SL_1 was the smooth L_1 loss function, which combines the advantages of L_1 and L_2 loss. They were defined as follows:

$$L_{adv}(x) = \tfrac{1}{2}\left(\left(D(x), G(x) - T_{real}\right)^2\right. \tag{3}$$

$$SL_1(x,y) = \begin{cases} \frac{1}{2}(y - G(x))^2, & |y - G(x)| < 1 \\ |y - G(x)| - 0.5, & otherwise \end{cases} \tag{4}$$

where T_{real}=1 and $T_{synthetic}$=0 were labels for the discriminant results of real and synthetic images, respectively. λ was the weight for SL_1 loss and is set to 10 in this study [31]. The proposed network was trained for 600 epochs with an initial learning rate of 0.0002, employing the Adam optimizer for optimization [32]. The network was implemented using PyTorch and executed on a NVIDIA GeForce RTX 4090 GPU.

Network training

The 3D cGAN was first pre-trained by 200 paired MC or analytical simulated NAC and CTAC DAT SPECT images to get the pre-trained model, respectively. Subsequently, 8, 24, and 80 pairs of NAC and CTAC clinical DAT SPECT images were employed to fine-tune the pre-trained MC simulation model (TLAC-MC) and analytical simulation model (TLAC-ANA). The FT process for TLAC-MC involved 3 different strategies: (i) TLAC-MC-st1: FT the first input layer i_1 and the last output layer o_1; (ii) TLAC-MC-st2: FT the first three down-sampling layers (i_1, i_2, i_3) and the last three up-sampling layers (o_1, o_2, o_3); (iii) TLAC-MC-st3: FT all layers ($n = 18$) L_{all}. Best FT strategy would be was used for TLAC-MC and TLAC-ANA later in the study for further comparison. The FT layers were depicted in Fig. 1a. Given the same number of available clinical data used for FT in TLAC, i.e., 8/24/80, various training strategies were also implemented for comparison, including training using limited amount (8/24/80) of clinical data (DLAC-CLI) as the baseline; with 6 times data augmentation ($8 \times 6/24 \times 6/80 \times 6$) using ± 10 degrees rotations and horizontal flip of the clinical data (DLAC-AUG); with a direct mixture (8/24/80 clinical + 200 MC simulation) of data (DLAC-MIX); and direct testing using the pre-trained MC model without FT (DLAC-MC). All training datasets were split into 7/8 for training and 1/8 validation. For testing, 20 clinical data were tested for models trained on 80 clinical datasets, while 50 clinical data were tested for models trained on 24 or 8 clinical datasets. All 100 clinical data were tested for DLAC-MC as a single fold. A 5-fold cross-validation was applied to training strategies utilizing 80 clinical datasets, while a 2-fold cross-validation was implemented for training strategies employing 24 or 8 clinical datasets to test all 100 clinical cases.

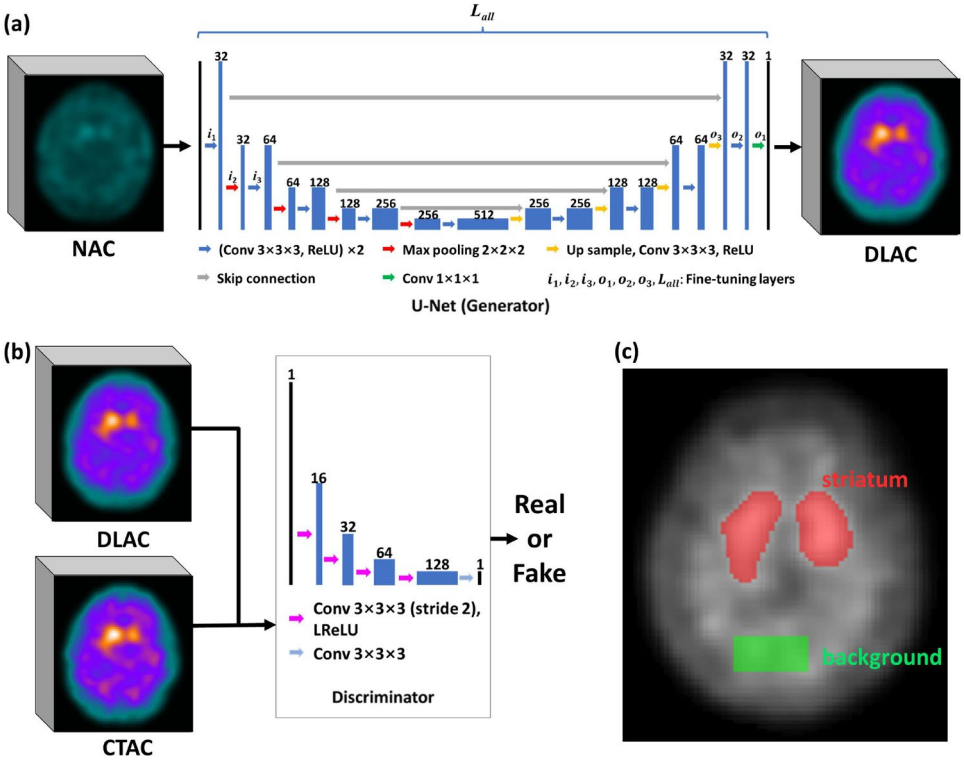

Fig. 1 The schematic diagram of the 3D cGAN architecture used in this study. (**a**) The U-Net generator and (**b**) the discriminator. The layers for fine-tuning are labeled on the U-Net generator. (**c**) Striatum (red mask) and background (green mask) regions-of-interest used in this study

Evaluation metrics

MC simulated CTAC images were compared with the clinical CTAC images using a voxel-based analysis of normalized mean square error (NMSE) (Eq. 5) and structural similarity index (SSIM) (Eq. 6) on the whole 3D brain region. Both indices were also assessed on clinical NAC, ChangAC, different DL-based AC images.

$$NMSE = \frac{\sum_{j=1}^{N}(x_j - \lambda_i)^2}{\sum_{j=1}^{N}\lambda_j^2} \tag{5}$$

$$SSIM = \frac{(2\mu_x\mu_\lambda + C_1)(2\sigma_{x,\lambda} + C_2)}{(\mu_x^2 + \mu_\lambda^2 + C_1)(\sigma_x^2 + \sigma_\lambda^2 + C_2)} \tag{6}$$

where N was the total number of voxels, λ as the voxel count value in the reference image, i.e., clinical CTAC image, x_j was the voxel count value in evaluated image, and j was the voxel index. μ_λ and σ_λ were the mean value and standard deviation of the reference image, μ_x and σ_x were the mean value and standard deviation of the evaluated image, $\sigma_{x,\lambda}$ was the cross-covariance between the evaluated image and the reference image. The constants $C_1 = (K_1 \times L)^2$ and $C_2 = (K_2 \times L)^2$ are used to stabilize the division when the denominator is very small. L represents the maximum voxel value between the 2 compared images. K_1 and K_2 are empirically set to 0.01 and 0.03 [33], respectively. Scatter plots of NMSE and SSIM results between reconstructed MC simulation and clinical data, as well as among clinical data were generated to evaluate the similarity between the two. A paired Mann-Whitney U test was used to evaluate their NMSE and SSIM differences, and was also employed to compare NMSE and SSIM results for different AC methods compared to their corresponding baseline method. For DL-based AC methods trained on different number of clinical datasets, their baselines were the DLAC-CLI models trained on the same number of clinical datasets. Meanwhile, NAC, ChangAC, and DLAC-MC were compared with DLAC-CLI trained on 8 clinical datasets. A p-value of <0.05 indicates a significant difference. Joint histogram analysis and linear regression were also performed to evaluate the distribution differences between various AC methods and their corresponding baselines.

For clinically relevant analysis, striatal binding ratio (SBR) (Eq. 3) was assessed for different AC images.

$$SBR = \frac{\bar{\lambda}_{st} - \bar{\lambda}_{bg}}{\bar{\lambda}_{bg}} \tag{7}$$

where $\bar{\lambda}_{st}$ was the mean count value in the striatum region and $\bar{\lambda}_{bg}$ was the mean count value in a uniform cerebellum region ($20 \times 10 \times 10$) (Fig. 1c). The striatum region was segmented using a threshold of 67% of the maximum intensity of the DAT SPECT images for each subject [34]. Bland-Altman plots were applied to SBR results to evaluate the differences among various AC methods compared to their corresponding baselines.

Results

Supplementary Fig. 1 presents sample MC simulated projections and reconstructed images with CTAC along with the clinical data, demonstrating a high degree of similarity. While Supplementary Fig. 2 further demonstrates the high similarity between simulation and clinical reconstructed images with CTAC, based on quantitative results.

Transfer learning strategies

Figure 2 displays sample images and error maps of various TL strategies for TLAC-MC with different amount of clinical data used for FT, while Table 1 presents the corresponding quantitative results. TLAC-MC-st3 consistently achieves the best result for both NMSE and SSIM among FT with 8, 24, and 80 clinical datasets. Sample images also show that TLAC-MC-st3 images are most similar to the CTAC image, followed by those of TLAC-MC-st2 and TLAC-MC-st1. As TLAC-MC-st3 shows the best performance among all FT strategies, it was used as the default TLAC-MC and TLAC-ANA method for subsequent evaluations.

AC methods evaluation

Figure 3 presents sample results of predicted AC images using DLAC-MC, DLAC-CLI, DLAC-AUG, DLAC-MIX, TLAC-MC, and TLAC-ANA methods, along with CTAC,

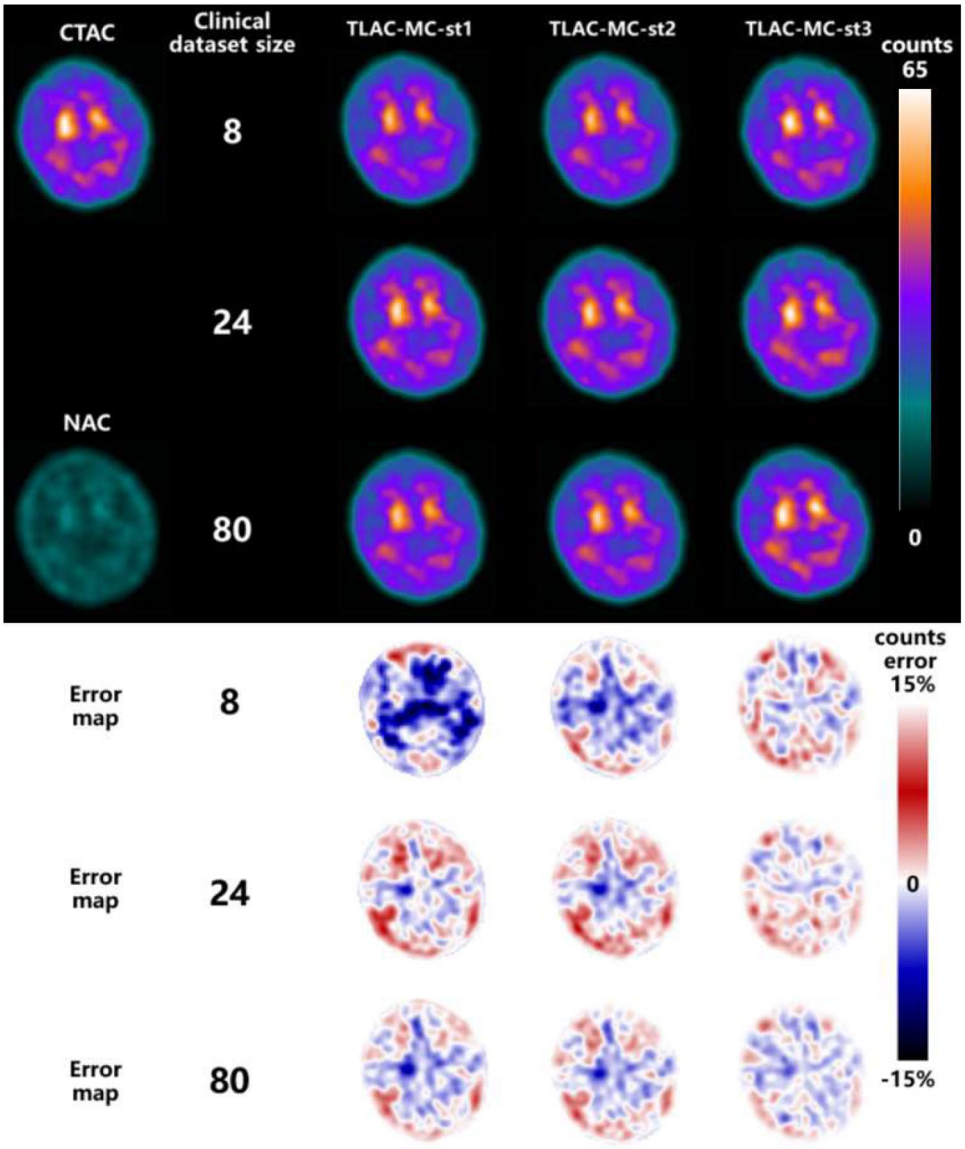

Fig. 2 Sample reconstructed images (top) and error maps (bottom) of TLAC methods with different TL strategies compared with CTAC and NAC

Table 1 NMSE and SSIM results (mean ± SD) of whole brain images for different TLAC strategies

TLAC strategy	Number of clinical datasets	NMSE	SSIM
TLAC-MC-st1	8	0.0192 ± 0.0107	0.9113 ± 0.0306
	24	0.0170 ± 0.0093	0.9179 ± 0.0296
	80	0.0140 ± 0.0105	0.9326 ± 0.0337
TLAC-MC-st2	8	0.0175 ± 0.0091	0.9181 ± 0.0313
	24	0.0158 ± 0.0089	0.9224 ± 0.0300
	80	0.0132 ± 0.0101	0.9331 ± 0.0341
TLAC-MC-st3	8	0.0143 ± 0.0082	0.9350 ± 0.0272
	24	0.0131 ± 0.0080	0.9351 ± 0.0275
	80	0.0125 ± 0.0100	0.9378 ± 0.0356

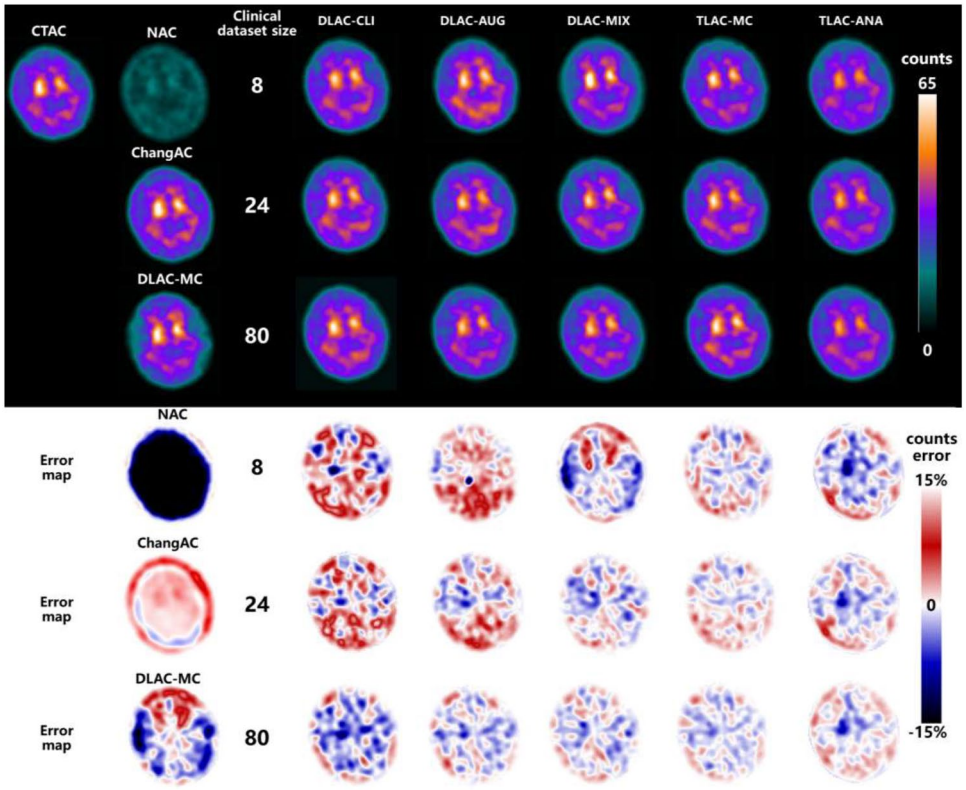

Fig. 3 Sample reconstructed images (top) and error maps (bottom) of different AC methods as compared to CTAC for a selected patient (age: 78, female)

NAC, and ChangAC images for comparison. The corresponding error maps are estimated using CTAC images as references. All AC methods show improved image quality compared to NAC. ChangAC overestimates the SPECT image counts values across most regions. Among the DL-based AC methods, TLAC-MC demonstrates the smallest errors across various number of clinical datasets. DLAC-AUG, DLAC-MIX, and TLAC-ANA also improve image quality compared to the conventional DLAC-CLI approach. Generally, all DL-based AC methods show less errors when the clinical data used for training increase.

NMSE and SSIM values of different AC methods compared to CTAC, as well as the training time for various DL-based AC methods are listed in Table 2. Consistent with visual assessment, TLAC-MC always achieves the best quantitative results among all methods, followed by DLAC-AUG. They all show significantly better results when

comparing with DLAC-CLI, except for the SSIM value with 80 clinical data used for training. TLAC-ANA performs better on NMSE but worse on SSIM when compared to both DLAC-MIX and DLAC-CLI across all number of clinical datasets. While DLAC-MIX performs significantly better on both SSIM and NMSE compares to DLAC-CLI when the clinical dataset size is 8 or 24. NAC, ChangAC, and DLAC-MC all perform significantly worse than DLAC-CLI trained using only 8 patient data. Increasing the number of clinical datasets improves DL-based AC performance and reduces the difference between DLAC-CLI and other DL-based AC methods. DLAC-CLI, TLAC-MC, and TLAC-ANA require similar training time for the same number of clinical datasets used, which is significantly shorter as compared to DLAC-AUG and DLAC-MIX.

Figure 4 depicts the joint histograms, their correlation coefficient (CC) and fitting R^2 values of various AC methods across all tested clinical data, with CTAC as the reference. Consistent with the former results, TLAC-MC always shows the highest correlation (CC = 0.9827 for 8 clinical data; CC = 0.9843 for 24 clinical data; CC = 0.9936 for 80 clinical data) with CTAC. DLAC-CLI (CC = 0.9527, R^2 = 0.9804 for 8 clinical data; CC = 0.9663, R^2 = 0.9776 for 24 clinical data; CC = 0.9870, R^2 = 0.9824 for 80 clinical data) shows higher correlation with CTAC than NAC (CC = 0.3640, R^2 = 0.8830), ChangAC (CC = 1.1526, R^2 = 0.9730) and DLAC-MC (CC = 0.9091, R^2 = 0.9700). For training with 8 and 24 clinical data, DLAC-AUG (CC = 0.9778, R^2 = 0.9835 for 8 clinical data; CC = 0.9806, R^2 = 0.9815 for 24 clinical data), DLAC-MIX (CC = 0.9728, R^2 = 0.9834 for 8 clinical data; CC = 0.9786, R^2 = 0.9840 for 24 clinical data) and TLAC-ANA (CC = 0.9695, R^2 = 0.9831 for 8 clinical data; CC = 0.9786, R^2 = 0.9858 for 24 clinical data) are all better than DLAC-CLI. For the use of 80 clinical datasets, the differences among different DL-based AC methods compared to CTAC become smaller.

Table 2 NMSE and SSIM results (mean ± std) of whole brain images for different AC methods

AC method	Number of clinical datasets	Training time	NMSE	*P*-value*	SSIM	*P*-value*
NAC	N/A	N/A	0.1428 ± 0.0287	< 0.0001	0.7275 ± 0.0382	< 0.0001
ChangAC			0.0309 ± 0.0123	< 0.0001	0.9060 ± 0.0292	0.0048
DLAC-MC		24 h 20 min	0.0304 ± 0.0112	< 0.0001	0.9127 ± 0.0306	0.2673
DLAC-AUG	8	4 h 7 min	0.0148 ± 0.0077	< 0.0001	0.9350 ± 0.0272	< 0.0001
	24	10 h 56 min	0.0141 ± 0.0065	< 0.0001	0.9351 ± 0.0275	< 0.0001
	80	36 h 7 min	0.0135 ± 0.0094	0.0048	0.9378 ± 0.0356	0.0637
DLAC-MIX	8	26 h 8 min	0.0166 ± 0.0090	0.0087	0.9293 ± 0.0194	0.0013
	24	28 h 4 min	0.0162 ± 0.0083	0.0359	0.9319 ± 0.0168	0.0040
	80	31 h 24 min	0.0145 ± 0.0082	0.3369	0.9354 ± 0.0153	0.1983
TLAC-MC	8	1 h 14 min	**0.0143 ± 0.0082**	< 0.0001	**0.9355 ± 0.0203**	< 0.0001
	24	2 h 23 min	**0.0131 ± 0.0080**	0.0028	**0.9359 ± 0.0187**	0.0002
	80	6 h 22 min	**0.0125 ± 0.0100**	< 0.0001	**0.9378 ± 0.0274**	0.2528
TLAC-ANA	8	1 h 15 min	0.0166 ± 0.0097	0.0211	0.9137 ± 0.0285	0.6130
	24	2 h 24 min	0.0155 ± 0.0067	0.0010	0.9163 ± 0.0262	0.0005
	80	6 h 22 min	0.0142 ± 0.0067	0.5268	0.9249 ± 0.0275	< 0.0001
DLAC-CLI	8	1 h 12 min	0.0202 ± 0.0110	N/A	0.9175 ± 0.0254	N/A
	24	2 h 27 min	0.0183 ± 0.0077	N/A	0.9247 ± 0.0229	N/A
	80	6 h 25 min	0.0150 ± 0.0073	N/A	0.9365 ± 0.0185	N/A

*Mann-Whitney U test was conducted between DLAC-CLI and other AC methods for the same number of clinical datasets. NAC, ChangAC, and DLAC-MC were compared with DLAC-CLI using 8 clinical data samples. Bolded values indicate the best results achieved among the different number of clinical datasets used for training/FT

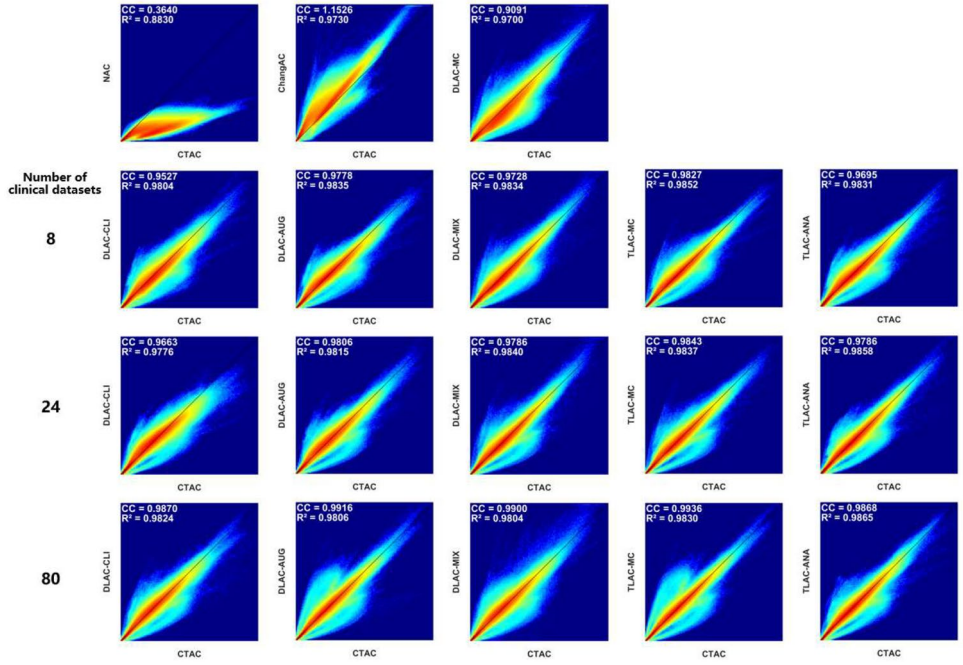

Fig. 4 Joint histogram and linear regression analysis of different AC methods across all 100 tested clinical data using CTAC as the reference

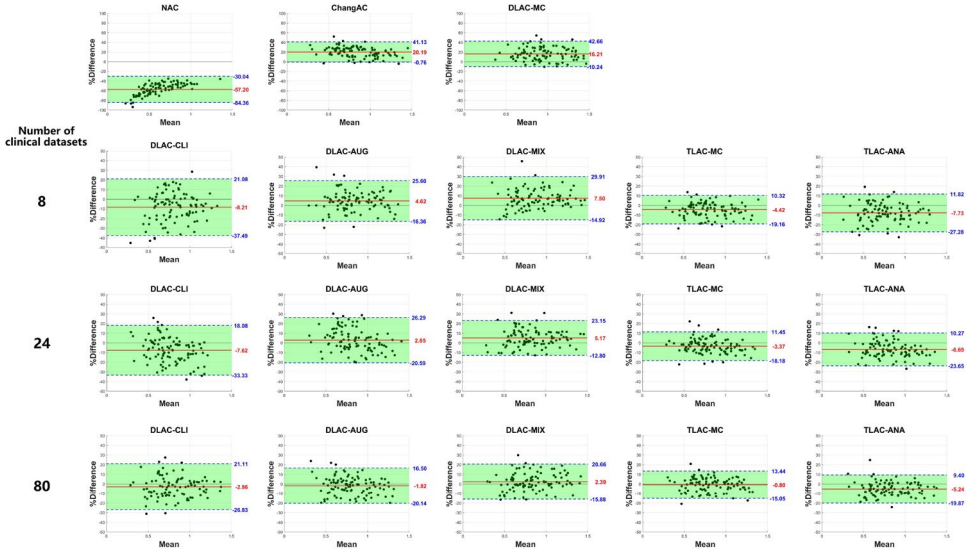

Fig. 5 Bland-Altman plots of SBR results of different AC methods across all 100 patients using CTAC as the reference

Figure 5 presents the Bland-Altman plots of SBR results for different AC methods across all 100 patients, using CTAC as the reference. NAC shows a substantial lower SBR value (mean %difference of -57.20%) compared to CTAC. Both ChangAC and DLAC-MC overestimate the SBR values, with mean %differences of 20.19% and 16.21%, respectively. DLAC-CLI, trained with 8, 24, and 80 clinical datasets, demonstrates better performance (mean SBR %differences of -8.21%, -7.62%, and-2.86%, respectively) compared to NAC, ChangAC, and DLAC-MC. DLAC-AUG, DLAC-MIX, and TLAC-MC all exhibit better performance to DLAC-CLI for the same number of clinical datasets used

in terms of both mean %differences and 95% confidence interval (CI) ranges. Among them, TLAC-MC shows both the smallest mean %difference (-4.42%, -3.37% and – 0.80% for 8, 24 and 80 clinical data respectively) and the smallest 95% CI of [-0.19.16, 10.32] for 8 clinical data, [-18.18, 11.45] for 24 clinical data, and [-15.05, 13.44] for 80 clinical data. Although TLAC-ANA shows a smaller 95% CI ([-27.28, 11.92] for 8, [-23.65, 10.27] for 24, and [-19.87, 9.40] for 80 clinical data), it shows a higher mean difference (-7.73 for 8 clinical data, -6.69 for 24 clinical data and – 5.24 for 80 clinical data) compares to DLAC-AUG and DLAC-MIX.

Discussions

In this study, we apply TL method for DAT SPECT AC using a model pre-trained purely by realistic simulation data, with limited clinical data for FT. Three different TL strategies are employed to fine-tune various layers of the U-Net generator. The results demonstrated that TLAC-MC-st3, which fine-tuned all layers, is the most effective strategy for this task. This result is consistent with a previous study on cardiac SPECT AC [35]. As suggested by Yosinski et al. [36], shallower layers in DL models are generally responsible for extracting domain-agnostic generals features, while deeper layers specialize in task-specific patterns. This finding aligns with our results and highlights the importance of FT all layers to maximize the TLAC performance.

Although we have shown that MC simulation data generated by our realistic phantom population demonstrate a high degree of similarity compared to clinical data based on both visual and quantitative evaluation (Supplementary Figs. 1 and 2), the difference between the two cannot be completely eliminated. Our results demonstrate that DLAC-MC, i.e., clinical data tested directly on the MC simulation-based model, performs worse than DLAC-CLI regardless of the number of clinical datasets used for training. However, after FT, TLAC-MC achieves substantial performance improvement as compared to DLAC-MC. This suggests that TL is an effective approach to bridge the gap between simulation and clinical datasets, enhancing the model's clinical feasibility and accuracy.

Compare to NAC and ChangAC, all DL-based AC methods in this work demonstrate significant performance improvement. As the performance of DL model demands a large number of training datasets, it becomes crucial to apply various data augmentation methods when the available training data is limited [37]. Our results indicate that both DLAC-AUG and DLAC-MIX could effectively improve the DL performance as compared to DLAC-CLI, particularly in scenarios with limited clinical data. On the other hand, our proposed TLAC-MC method not only outperforms other DL-based AC methods, it also greatly reduces the training time. This finding demonstrates that TLAC-MC is an effective approach for clinical utilization, addressing both data and computational demand.

Haimour et al. [38] demonstrated that models pre-trained on a large datasets of medical images outperformed those pre-trained on natural images for medical image translation tasks, when FT using medical images are applied on the pre-trained models. Their findings suggest that the similarity between the data used for the pre-trained network and the target test data may greatly influence the performance of a TL model. In this work, we also evaluate the performance difference between TL models fine-tuned using two pre-trained models: one trained on realistic MC simulation data and another trained on more accessible but less realistic analytical simulation data. The results indicate that

TLAC-ANA performs worse than TLAC-MC, as expected, but still outperforms the baseline model DLAC-CLI when same number of clinical data are used in most evaluations. This finding reinforces the idea that the closer the pre-training data are to the target test data, the better the pre-trained model will perform. In this work, a realistic phantom population incorporating clinically relevant anatomical variations and tracer distributions was utilized to train pre-trained models, yielding promising results. Notably, even less realistic analytical simulation data demonstrated the ability to enhance DL performance. However, the trade-off between the realism of the simulation data and the resulting model performance warrants further investigation.

Our experimental results also show that for clinical evaluation, NAC is difficult to distinguish the striatum and background regions. The SBR values from NAC were underestimated as compared to the reference CTAC. In contrast, all AC methods effectively improve the differentiation between the striatum and background in DAT SPECT, yielding SBR results closer to those of the CTAC reference. Notably, TLAC-MC exhibits minimal SBR differences compared to CTAC, even when fine-tuned with only 8 clinical data, showing its robustness in clinical evaluations.

One limitation of this work is that the pre-trained model for TLAC is trained using only 200 simulation data. Future studies could explore the impact of varying the number of simulation data for pre-training. Additionally, this study utilizes only the U-Net architecture as the generator for the pre-trained model, evaluating more advanced network architectures could potentially further enhance the TLAC performance. Furthermore, only data from one activity distribution, i.e., ^{99m}Tc-TRODAT-1, and one scanner, i.e., Siemens Symbia, are employed in this study. Our proposed method should be applicable to other DAT SPECT/PET with different radiopharmaceuticals. In fact, another phantom population for ^{123}I-ioflupane DAT SPECT has already been established [28]. Further investigation using data with other DAT tracers to further validate our proposed method is warranted.

Conclusions

TLAC is feasible for DAT SPECT with a pre-trained model generated based on realistic simulation data. Both qualitative and quantitative results show that the proposed TLAC-MC and TLAC-ANA outperform DLAC-CLI. While TLAC-MC achieves the best performance compared to other DL-based AC methods such as TLAC-ANA, DLAC-AUG and DLAC-MIX, particularly when clinical datasets are limited.

Abbreviations

AC	Attenuation correction
CC	Correlation coefficient
cGAN	Conditional generative adversarial network
CI	Confidence interval
ChangAC	Chang's attenuation correction method
D	Discriminator
DAT	Dopamine transporter
DEW	Double energy window
DL	Deep learning
DLAC-AUG	Deep learning-based attenuation correction with clinical data augmentation for training
DLAC-CLI	Deep learning-based attenuation correction with clinical data for training
DLAC-MC	Deep learning-based attenuation correction with Monte-Carlo simulation data for training
DLAC-MIX	Deep learning-based attenuation correction with mixed Monte Carlo simulation data and clinical data for training
FOV	Field-of-view
FT	Fine-tuning
G	Generator

MC	Monte Carlo
NMSE	Normalized mean square error
PD	Parkinson's disease
ReLU	Rectified linear unit
SBR	Striatal binding ratio
SD	Standard deviation
SPECT	Single photon emission computed tomography
SSIM	Structural similarity index
TL	Transfer learning
TLAC	Transfer learning-based attenuation correction
TLAC-MC	Transfer learning-based attenuation correction with Monte Carlo simulation data for fine-tuning
TLAC-ANA	Transfer learning-based attenuation correction with analytical simulation data for fine-tuning
VI	Visual Interpretation

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40658-025-00756-1.

Supplementary Material 1

Acknowledgements
The computing was performed in part at SICC which is supported by SKL-IOTSC, University of Macau.

Author contributions
WB.H contributed to the study design and implementation, data preprocess and analysis, drafting the manuscript, and preparing the tables and figures. H.J contributed to the manuscript editing, revision and data curation. Y.D, HY.W and H.S contributed to manuscript editing, revision and methodology. GU.H contributed to data acquisition and clinical interpretation. G.M contributed to the study design, data analysis, curation, interpretation, supervision, manuscript writing and editing. All authors critically read and provided feedback on previous versions of the manuscript. All authors discussed the results and implications and commented on the manuscript.

Funding
This work was supported by Science and Technology Development Fund, Macau (0178/2024/AMJ and 0016/2023/RIB1), University of Macau and University of Macau Development Foundation (MYRG-GRG2024-00061-FST-UMDF).

Data availability
Authors will share data upon request to the corresponding author.

Declarations

Ethics approval and consent to participate
The patients data used in this study are under local ethics approval (SCMH_IRB No: 1110704).

Consent for publication
All authors are consent for publication.

Competing interests
The authors declare that they have no competing interests.

Received: 22 January 2025 / Accepted: 18 April 2025
Published online: 06 May 2025

References

1. Feigin VL, Nichols E, Alam T, Bannick MS, Beghi E, Blake N, et al. Global, regional, and National burden of neurological disorders, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurol. 2019;18:459–80.
2. Palermo G, Ceravolo R. Molecular imaging of the dopamine transporter. Cells. 2019;8:872.
3. Bega D, Kuo PH, Chalkidou A, Grzeda MT, Macmillan T, Brand C, et al. Clinical utility of DaTscan in patients with suspected parkinsonian syndrome: a systematic review and meta-analysis. Npj Parkinson's Disease. 2021;7:43.
4. Wu H, Lou C, Huang Z, Shi G. SPECT imaging of dopamine transporters with 99mTc-TRODAT-1 in major depression and Parkinson's disease. J Neuropsychiatry Clin Neurosci. 2011;23:63–7.
5. Jaszczak RJ, Coleman RE, Whitehead FR. Physical factors affecting quantitative measurements using camera-based single photon emission computed tomography (SPECT). IEEE Trans Nucl Sci. 1981;28:69–80.
6. Hayashi M, Deguchi J, Utsunomiya K, Yamada M, Komori T, Takeuchi M, et al. Comparison of methods of Attenuation and scatter correction in brain perfusion SPECT. J Nucl Med Technol. 2005;33:224–9.
7. Bieńkiewicz M, Górska-Chrząstek M, Siennicki J, Gajos A, Bogucki A, Mochecka-Thoelke A, et al. Impact of CT based Attenuation correction on quantitative assessment of DaTSCAN (123 I-Ioflupane) imaging in diagnosis of extrapyramidal diseases. Nuclear Med Rev. 2008;11:53–8.
8. Zaidi H, Hasegawa B. Determination of the Attenuation map in emission tomography. J Nucl Med. 2003;44:291–315.
9. Technavio. Global SPECT market 2017–2021 Technical Report. https://technavio.com/report/global-medical-imaging-global-spect-market-2017-2021

10. Boas FE, Fleischmann D. CT artifacts: causes and reduction techniques. Imaging Med. 2012;4:229–40.
11. Warwick J, Rubow S, du Toit M, Beetge E, Carey P, Dupont P. The role of CT-Based Attenuation correction and collimator blurring correction in striatal spect quantification. Int J Mol Imaging. 2011;2011:195037.
12. Lee Y-W, Yang C-C, Mok GS, Wu T-H. Infant cardiac CT angiography with 64-slice and 256-slice CT: comparison of radiation dose and image quality using a pediatric Phantom. PLoS ONE. 2012;7:e49609.
13. Chang L-T. A method for Attenuation correction in radionuclide computed tomography. IEEE Trans Nucl Sci. 1978;25:638–43.
14. Du Y, Jiang H, Lin C-N, Peng Z, Sun J, Chiu P-Y et al. Generative adversarial network-based Attenuation correction for 99mTc-TRODAT-1 brain SPECT. Front Med. 2023;10.
15. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data. 2016;3:1–40.
16. Sun H, Du Y, Lin C-N, Jiang H, Huang W, Chiu P-Y, et al. Cross-Tracer and Cross-Scanner transfer Learning-Based Attenuation correction for brain SPECT. IEEE Transactions on Radiation and Plasma Medical Sciences; 2024.
17. Segars W, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT Phantom for multimodality imaging research. Med Phys. 2010;37:4902–15.
18. Buvat I, Castiglioni I. Monte Carlo simulations in SPET and PET. QJ Nucl Med. 2002;46:48–61.
19. Ljungberg M, Strand S-E. A Monte Carlo program for the simulation of scintillation camera characteristics. Comput Methods Programs Biomed. 1989;29:257–72.
20. Jan S, Santin G, Strul D, Staelens S, Assié K, Autret D, et al. GATE: a simulation toolkit for PET and SPECT. Phys Med Biol. 2004;49:4543.
21. Huang W, Mok GS. Multi-pinhole collimator design in different numbers of projections for brain SPECT. Front Med. 2023;10.
22. Mok GS, Du Y, Wang Y, Frey EC, Tsui BM. Development and validation of a Monte Carlo simulation tool for multi-pinhole SPECT. Mol Imaging Biology. 2010;12:295–304.
23. Segars W, Bond J, Frush J, Hon S, Eckersley C, Williams CH, et al. Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization. Med Phys. 2013;40:043701.
24. Abadi E, Segars WP, Tsui BM, Kinahan PE, Bottenus N, Frangi AF, et al. Virtual clinical trials in medical imaging: a review. J Med Imaging. 2020;7:042805.
25. Djang DS, Janssen MJ, Bohnen N, Booij J, Henderson TA, Herholz K, et al. SNM practice guideline for dopamine transporter imaging with 123I-ioflupane SPECT 1.0. J Nucl Med. 2012;53:154–63.
26. Leube J, Gustafsson J, Lassmann M, Salas-Ramirez M, Tran-Gia J. Analysis of a deep learning-based method for generation of SPECT projections based on a large Monte Carlo simulated dataset. EJNMMI Phys. 2022;9:47.
27. Patton JA, Turkington TG. SPECT/CT physical principles and Attenuation correction. J Nucl Med Technol. 2008;36:1–10.
28. Huang W, Mok GS. Development of a Digital Brain Phantom Population for Radionuclide Imaging Research in Parkinson's Disease. 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC): IEEE; 2022. pp. 1–3.
29. Segars W, Mok GS, Tsui BM. Investigation of respiratory gating in quantitative myocardial SPECT. IEEE Trans Nucl Sci. 2009;56:91–6.
30. Frey E, Tsui B. A practical projector-backprojector modeling attenuation, detector response, and scatter for accurate scatter compensation in SPECT. Conference Record of the 1991 IEEE nuclear science symposium and medical imaging conference: IEEE; 1991. pp. 1777-81.
31. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. pp. 1125-34.
32. Kingma DP. Adam: A method for stochastic optimization. ArXiv Preprint arXiv:14126980. 2014.
33. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13:600–12.
34. Fang Y-HD, Chiu S-C, Lu C-S, Yen T-C, Weng Y-H. Fully automated quantification of the striatal uptake ratio of [99mTc]-TRODAT with SPECT imaging: evaluation of the diagnostic performance in Parkinson's disease and the Temporal regression of striatal tracer uptake. Biomed Res Int. 2015;2015:461625.
35. Chen X, Pretorius PH, Zhou B, Liu H, Johnson K, Liu Y-H, et al. Cross-vender, cross-tracer, and cross-protocol deep transfer learning for Attenuation map generation of cardiac SPECT. J Nuclear Cardiol. 2022;29:3379–91.
36. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? Adv Neural Inf Process Syst. 2014;27.
37. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6:1–48.
38. Haimour F, Al-Sayyed R, Mahafza W, Al-Kadi S. O. Bidirectional brain image translation using transfer learning from generic pre-trained models. Comput Vis Image Underst. 2024;248.

Publisher's note

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