European Journal of Nuclear Medicine and Molecular Imaging (2019) 46:2800–2811
https://doi.org/10.1007/s00259-019-04502-5

ORIGINAL ARTICLE

Automatic classification of dopamine transporter SPECT: deep convolutional neural networks can be trained to be robust with respect to variable image characteristics

Markus Wenzel[1] · Fausto Milletari[2,3] · Julia Krüger[4] · Catharina Lange[5] · Michael Schenk[6] · Ivayla Apostolova[6] · Susanne Klutmann[6] · Marcus Ehrenburg[7] · Ralph Buchert[6]

Received: 3 January 2019 / Accepted: 22 August 2019 / Published online: 31 August 2019

Abstract

Purpose This study investigated the potential of deep convolutional neural networks (CNN) for automatic classification of FP-CIT SPECT in multi-site or multi-camera settings with variable image characteristics.

Methods The study included FP-CIT SPECT of 645 subjects from the Parkinson's Progression Marker Initiative (PPMI), 207 healthy controls, and 438 Parkinson's disease patients. SPECT images were smoothed with an isotropic 18-mm Gaussian kernel resulting in 3 different PPMI settings: (i) original (unsmoothed), (ii) smoothed, and (iii) mixed setting comprising all original and all smoothed images. A deep CNN with 2,872,642 parameters was trained, validated, and tested separately for each setting using 10 random splits with 60/20/20% allocation to training/validation/test sample. The putaminal specific binding ratio (SBR) was computed using a standard anatomical ROI predefined in MNI space (AAL atlas) or using the hottest voxels (HV) analysis. Both SBR measures were trained (ROC analysis, Youden criterion) using the same random splits as for the CNN. CNN and SBR trained in the mixed PPMI setting were also tested in an independent sample from clinical routine patient care (149 with non-neurodegenerative and 149 with neurodegenerative parkinsonian syndrome).

Results Both SBR measures performed worse in the mixed PPMI setting compared to the pure PPMI settings (e.g., AAL-SBR accuracy = 0.900 ± 0.029 in the mixed setting versus 0.957 ± 0.017 and 0.952 ± 0.015 in original and smoothed setting, both $p < 0.01$). In contrast, the CNN showed similar accuracy in all PPMI settings (0.967 ± 0.018, 0.972 ± 0.014, and 0.955 ± 0.009 in mixed, original, and smoothed setting). Similar results were obtained in the clinical sample. After training in the mixed PPMI setting, only the CNN provided acceptable performance in the clinical sample.

Conclusions These findings provide proof of concept that a deep CNN can be trained to be robust with respect to variable site-, camera-, or scan-specific image characteristics without a large loss of diagnostic accuracy compared with mono-site/mono-camera settings. We hypothesize that a single CNN can be used to support the interpretation of FP-CIT SPECT at many different

This article is part of the Topical Collection on Advanced Image Analyses (Radiomics and Artificial Intelligence)

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00259-019-04502-5) contains supplementary material, which is available to authorized users.

✉ Ralph Buchert
r.buchert@uke.de

[1] Fraunhofer Institute for Medical Image Computing MEVIS, Bremen, Germany

[2] Computer Aided Medical Procedures & Augmented Reality, Technical University Munich, München, Germany

[3] Nvidia, Los Angeles, USA

[4] Jung diagnostics, Hamburg, Germany

[5] Department of Nuclear Medicine, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[6] Department for Diagnostic and Interventional Radiology and Nuclear Medicine, University Hospital Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany

[7] Pinax Pharma, Bad Liebenwerda, Germany

sites using different acquisition hardware and/or reconstruction software with only minor harmonization of acquisition and reconstruction protocols.

Keywords Deep learning · Convolutional neural network · Dopamine transporter · SPECT · FP-CIT · Domain adaption

Introduction

SPECT with N-ω-fluoropropyl-2β-carbomethoxy-3β-(4-I-123-iodophenyl)nortropane (FP-CIT) is widely used for detection (or exclusion) of nigrostriatal degeneration in clinically uncertain parkinsonian syndromes (PS) [1–4]. The interpretation of FP-CIT SPECT is primarily based on visual reading which provides high accuracy with respect to the differentiation between neurodegenerative PS (with nigrostriatal degeneration) and non-neurodegenerative PS (without nigrostriatal degeneration) [5, 6].

Visual reading can be complemented by semi-quantitative analysis of striatal FP-CIT uptake [7, 8], the specific binding ratio (SBR) being still the most widely used semi-quantitative measure in FP-CIT SPECT [9–12]. A recent study by Booij and co-workers found that the combination of visual reading and SBR analysis was not inferior to visual reading alone and resulted in increased reader confidence [7]. There was also some weak evidence that adding SBR analysis might allow physicians with only a little experience in the interpretation of FP-CIT SPECT to reach similar performance as experienced readers [7]. Albert and co-workers re-analyzed FP-CIT scans that had initially been interpreted as inconclusive and found that SBR analysis with correction for patient age and camera-specific variability might help to clarify some of these cases [13]. Together, these findings suggest that the utility of the conventional SBR is limited in clinical routine except for some possible incremental value for inexperienced readers and in borderline cases.

The major factor limiting the utility of SBR analysis is its sensitivity with respect to the site- and/or camera-specific variability of SPECT image characteristics caused by differences in acquisition and reconstruction protocols [14–21]. This sensitivity complicates the sharing of normal databases and SBR cut-off values between sites/cameras. In addition, the SBR is not only affected by site-/camera-specific variability but also by scan-specific variabilities such as head motion and varying radius of rotation of the camera heads [22, 23]. This results in increased variability of the SBR reducing its power to detect actual changes of striatal DAT availability also in mono-site/mono-camera settings.

Fully automatic classification of FP-CIT SPECT images may support visual reading and improve reader confidence similar to or better than SBR analysis. The rationale for this is that high-dimensional expert systems might be trained to account for site-, camera-, and scan-specific variability of image characteristics of FP-CIT SPECT, similar to human readers that are usually less sensitive to this variability than conventional SBR analysis (Fig. 1). The aim of this study, therefore, was to assess the performance of a deep learning-based classifier (convolutional neural network, CNN) [24, 25] to deal with varying image characteristics in FP-CIT SPECT.

Other deep learning applications in nuclear medicine brain imaging include image classification of brain FDG PET in cognitively impaired patients [26–32], detection, delineation, and classification of brain tumors in brain PET with FDG or amino acid tracers [33, 34], image registration and spatial normalization of brain amyloid PET [35], image enhancement/denoising in brain FDG PET [36, 37], estimating synthetic brain FDG PET images from individual MRI [38], and generating pseudo-CT images from individual MRI for attenuation correction in brain PET [39].

Materials and methods

PPMI sample

The first sample of FP-CIT SPECT images used in this study was obtained from the Parkinson's Progression Markers Initiative (PPMI) (www.ppmi-info.org/data) [40]. Up-to-date information on the PPMI is given at www.ppmi-info.org. The PPMI is a longitudinal, multi-center study that aims to assess the progression of clinical features, imaging, and biologic markers in patients with Parkinson's disease (PD) and healthy control (HC) subjects. Details of the PPMI eligibility criteria are given at http://www.ppmi-info.org/wp-content/uploads/2014/01/PPMI-AM7-Protocol.pdf. Details of the PPMI FP-CIT SPECT protocol are given at http://www.ppmi-info.org/study-design/research-documents-and-sops/ [40]. Raw projection data had been transferred to the PPMI imaging core lab for central image reconstruction using an iterative ordered-subsets-expectation-maximization algorithm on a HERMES workstation.

We downloaded the first FP-CIT scan of all HC subjects and all PD patients (November 22, 2017). The visual inspection resulted in the exclusion of 3 HC scans because of reduced striatal FP-CIT uptake (PPMI-ID 3221, 3478, and 4095) and 8 PD scans because of normal striatal FP-CIT uptake (3027, 3289, 3290, 3534, 3618, 3623, 3660, and 3863). A total of 645 FP-CIT SPECTs were included, 207 of HC subjects and 438 of PD patients.

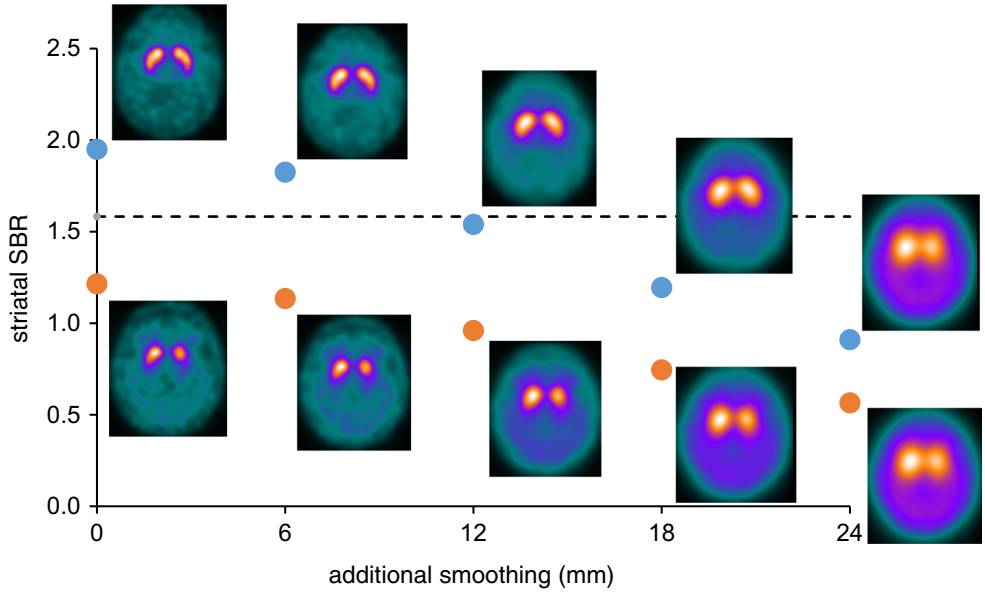

Fig. 1 Specific binding ratio (SBR) of the whole striatum (hottest voxels analysis as described in the "Materials and methods" section, mean of both hemispheres) as a function of the amount of additional smoothing achieved by post-filtering the original unsmoothed image with a 3-dimensional Gaussian filter to simulate site-specific variability of spatial resolution. This was done for a healthy control subject (upper curve) and a Parkinson's disease subject (lower curve) of the Parkinson's Disease Progression Marker Initiative (subject-ID 3013, 3014). A slab view through the striatum of the (smoothed) image is shown with each data point. The color table was scaled to the dynamic range of voxel intensities separately for each image. The dashed line indicates the cut-off to detect Parkinson's disease by SBR analysis of original (unsmoothed) images. The performance of this fixed cut-off worsens with decreasing spatial resolution. In particular, it results in misclassification of normal scans as Parkinson's disease when applied to images with 12 mm or more post-smoothing. However, correct visual interpretation of the images is possible until at least 18-mm post-smoothing

In order to simulate FP-CIT SPECT with very low spatial resolution, we applied an isotropic Gaussian filter with 18-mm full-width-at-half-maximum to each of the original 3-dimensional PPMI SPECT images prior to any further processing. This resulted in 3 different PPMI settings: (i) original (unsmoothed) (207 HC, 438 PD), (ii) smoothed (207 HC, 438 PD), and (iii) mixed setting comprising all original and all smoothed images (414 HC, 876 PD).

Clinical sample

Two hundred ninety-eight patients from routine clinical patient care were recruited retrospectively from the database of the University Medical Center Hamburg-Eppendorf. The patients were categorized into "neurodegenerative PS" and "non-neurodegenerative PS." The neurodegenerative group ($n = 149$, 46.3% females, 64.9 ± 10.7 y) comprised (i) the Lewy body disease spectrum ($n = 132$, 88.6%) including PD, PD dementia, and dementia with Lewy bodies, and (ii) atypical Parkinsonian syndromes ($n = 17$, 11.4%) including multiple systems atrophy, progressive supranuclear palsy, and corticobasal degeneration. The non-neurodegenerative group ($n = 149$, 50.3% females, 66.2 ± 11.8 y) comprised essential tremor, drug-induced parkinsonism, several types of dystonia, psychogenic parkinsonism, and various other diagnoses not associated with nigrostriatal degeneration. The clinical diagnosis as standard of truth was taken from the written report of a movement disorder specialist in the patient's file at least 12 months after FP-CIT SPECT in all 149 patients with neurodegenerative PS (mean follow-up 41 ± 23 months, range 13–96 months) and in 44 of the patients with non-neurodegenerative PS (38 ± 22 months, 13–96 months). The remaining 105 patients with non-neurodegenerative PS had less than 12 months follow-up and were included to achieve class balance.

FP-CIT SPECT had been performed according to common guidelines [41] with a double-head SPECT system (Siemens Symbia T2 or Siemens E.CAM). In order to ensure consistent image reconstruction in all patients, projection data were retrieved from the archive and reconstructed retrospectively. Two different reconstruction algorithms were used in each patient. First, SPECT images were reconstructed by filtered backprojection implemented in the SPECT system software (Butterworth filter of 5th order with cut-off 0.6 cycles/pixel). Uniform post-reconstruction attenuation correction was performed according to Chang ($\mu = 0.12$/cm), no scatter correction. Second, SPECT images were reconstructed using the ordered-subsets-expectation-maximization algorithm with resolution recovery implemented in the HybridRecon-Neurology tool of the Hermes SMART workstation v1.6 with parameter settings recommended for FP-CIT SPECT by

Hermes (effective number of iterations 80, post-filtering with 3-dimensional Gaussian kernel of 7-mm full-width-at-half-maximum, uniform attenuation correction with narrow-beam attenuation coefficient 0.146/cm, simulation-based scatter correction, resolution recovery with a Gaussian model). Representative images reconstructed with filtered backprojection and iterative reconstruction are shown in Fig. 2.

The FP-CIT SPECT data of the clinical sample are available from the corresponding author on request.

Semi-quantitative SBR analysis

Individual SPECTs were normalized (affine) to a custom-made FP-CIT template (Fig. 3) in MNI space using SPM12 [42]. Voxel intensity was scaled voxel-wise to the 75th percentile of the voxel intensities in a reference region comprising whole-brain except striata, thalamus, brain stem, and ventricles [43, 44] (Fig. 3).

Anatomical ROIs for unilateral caudate and putamen predefined in MNI space by the Automatic Anatomical Labeling atlas (AAL) were applied to the voxel-wise scaled image [45] (Fig. 3). The mean value of scaled voxel intensities in the AAL ROIs was used to calculate the conventional SBR (AAL-SBR = mean ROI value − 1).

In addition, hottest voxels (HV) analysis was performed, separately for caudate and putamen using large unilateral ROIs predefined in MNI space (Fig. 3). The union of caudate and putamen ROI was used to compute the SBR of the entire striatum. The ROIs for HV analysis was much bigger than the actual striatal volume in order to guarantee that all striatal counts were included. The number of hottest voxels to be averaged was fixed to a total volume of 5, 10, and 15 ml for unilateral caudate, putamen, and whole striatum, respectively (HV-SBR).

In the PPMI sample, SBR values provided at the PPMI homepage were used for comparison (PPMI-SBR, StudyData/Imaging/DaTSCAN/DaTscan_Analysis.csv).

For each SBR measure (AAL, HV, and PPMI) and each ROI, the minimum over both hemispheres was used in all analyses.

The diagnostic power of a given SBR measure was assessed as follows. First, the "optimal" cut-off was determined by the Youden criterion [46] applied to the receiver operating characteristic curve of the SBR for differentiation of PD versus HC (PPMI sample) or neurodegenerative PS versus non-neurodegenerative PS (clinical sample) in the training set (subsection "Performance testing in PPMI sample"). The cut-off determined in the training set was then applied for classification of the images in the test set. Overall accuracy, sensitivity, and specificity in the test set were used as performance measures.

CNN methods

We contrast the semi-quantitative SBR analysis with a deep neural network-based approach [24]. For image analysis, a variant of deep neural networks, the convolutional neural networks (CNN) have a reputation for solving cognitive tasks on par with or even better than human observers. The fundamental working principle is that the task is solved end-to-end by the CNN, which means that no human knowledge (like activity ratios and their cut-off) are engineered into the process, but the CNN itself learns the salient characteristics of the data based on a sufficiently large number of examples (input) and their class label (output). This is the reason why deep neural network-based methods run under the term of "feature learning" approaches, as opposed to "feature engineering" approaches like semi-quantitative SBR analysis. In the past several years, deep neural network-based methods have

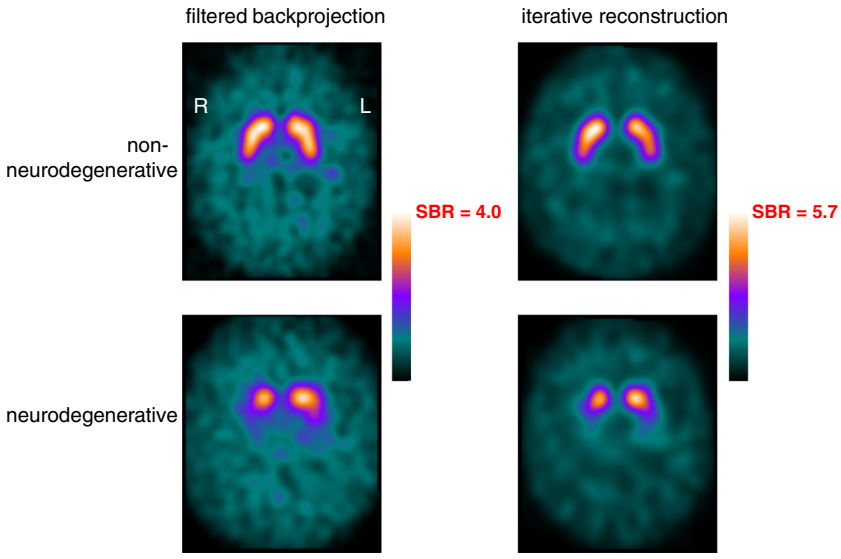

Fig. 2 FP-CIT SPECT images of a patient with non-neurodegenerative PS (upper row) and a patient with neurodegenerative PS (lower row) from the clinical sample reconstructed with filtered backprojection (left column) or iterative reconstruction (right column). The upper threshold of the color table strongly differs between filtered backprojection and iterative reconstruction

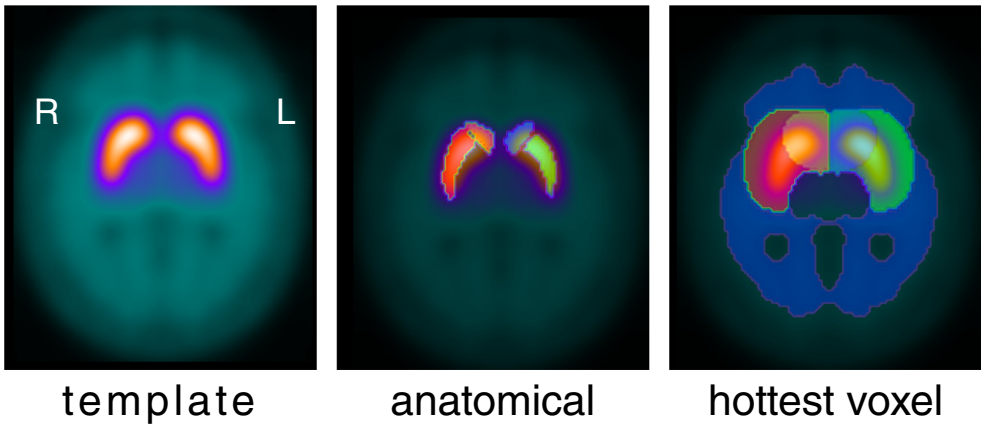

Fig. 3 Custom-made FP-CIT template (left), anatomical caudate and putamen ROIs from the Automatic Anatomic Labeling atlas used for conventional SBR analysis (middle), large ROIs of caudate and putamen for hottest voxels analysis (right), and reference region to estimate non-displaceable FP-CIT binding (right)

helped solving all kinds of medical problems and have exhibited excellent performance in medical imaging tasks [25].

For the task of classifying FP-CIT images, we developed a custom CNN architecture composed of 2D convolutional layers, batch normalization layers for regularization, and dense layers to convert the output of the convolutional layers into the final classification output. The model was trained against the categorical cross-entropy loss using the Adam optimizer with default parameters of the Keras implementation. The detailed structure of CNN is given in Fig. 4. The total number of CNN parameters was 2,872,642 (2,871,234 trainable, 1408 non-trainable).

The CNN was trained (≤ 500 iterations), validated, and tested on 2-dimensional FP-CIT slabs through the striatum ($2 \times 2 \times 12$ mm^3). Representative slabs are shown in Fig. 1. The best performing model according to validation performance during training was kept for testing.

Performance testing in PPMI sample

To avoid overfitting, the PPMI sample was randomly split into training, validation, and test sample (60/20/20%). The PD group was undersampled to achieve class balance (HC/PD): training sample 125/125, validation sample 41/41, and test sample 41/41. The same splits were used for original and smoothed PPMI setting. They were also used for the mixed PPMI set by randomly selecting either the original or the smoothed image of a given subject in order to avoid bias by including the smoothed image of a subject in the test set whose unsmoothed image was in the training set (or vice versa).

The CNN was also tested in two cross-site settings. For each split, the model trained in the original training sample was evaluated in the smoothed test sample, and vice versa.

A total of 10 random splits was used. The same splits were used for the SBR measures and the CNN. Performance measures (accuracy, sensitivity, and specificity) were averaged over the 10 splits after removing outliers (values below first quartile − 1.5 × interquartile range or above third quartile + 1.5 × interquartile range). Overall accuracy was compared between methods (PPMI-SBR, AAL-SBR, HV-SBR, CNN) and settings (original, smoothed, mixed) using the signed-rank Wilcoxon test. All tests were performed two-sided.

Application of models trained in the PPMI sample to the clinical sample

The mean HV-SBR cut-offs obtained in the different PPMI settings were applied to the clinical sample. The CNN was trained and validated in an 85/15% random split of the mixed PPMI sample (unbalanced with respect to class and including

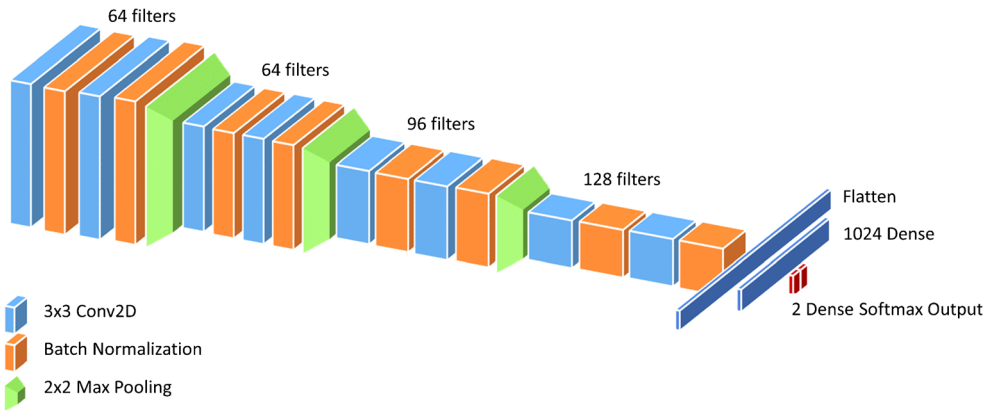

Fig. 4 Structure of the CNN used in the analyses presented here. The total number of CNN parameters was 2,872,642 (2,871,234 trainable, 1408 non-trainable)

both original and smoothed image of the same subject) and then tested in the clinical sample.

Performance testing in the clinical sample

The clinical sample was randomly split into training, validation, and test sample (60/20/20%). The same splits were used for filtered backprojection and iterative reconstruction. The same splits were also used for the mixed setting by randomly selecting either the filtered backprojected image or the iteratively reconstructed image. A total of 10 random splits was used. The same splits were used for AAL-SBR, HV-SBR, and the CNN.

Performance measures were averaged over the 10 random splits after removing outliers as before. Overall accuracy was compared between settings (filtered backprojection, iterative reconstruction, mixed) using the signed-rank Wilcoxon test. All tests were performed two-sided.

Results

Classification performance of the semi-quantitative SBR measures and the CNN in the original PPMI setting is summarized in Table 1. None of the differences in overall classification accuracy between PPMI-SBR, AAL-SBR, HV-SBR, and CNN reached statistical significance. Caudate and whole striatum SBR provided considerably worse performance compared with the putamen SBR (supplementary Table 1) and, therefore, are not discussed any further here.

Classification performance in pre-smoothed and mixed PPMI setting is given in Table 2. The accuracy of the SBR measures did not differ between the original and the smoothed PPMI setting (AAL-SBR: $p = 0.500$; HV-SBR: $p = 0.125$). However, the performance was worse in the mixed setting for both SBR measures. The accuracy reduction of the AAL-SBR in the mixed setting was statistically significant compared with the original ($p = 0.004$) and compared with the smoothed setting ($p = 0.006$). The reduction of HV-SBR accuracy in the mixed setting was statistical significant compared with the original setting ($p = 0.027$). By contrast, the accuracy of the CNN did not differ between the mixed PPMI setting and any of the pure PPMI settings. CNN accuracy was even slightly larger in the mixed setting than in the smoothed setting, but the difference did not reach statistical significance ($p = 0.188$).

CNN performance in the cross-site PPMI settings was as follows (trained in original and tested in pre-smoothed/trained in pre-smoothed and tested in original): accuracy = 0.629 ± 0.069/0.945 ± 0.058, sensitivity = 1.000/0.978 ± 0.014, and specificity = 0.259 ± 0.139/0.912 ± 0.121.

Classification performance of models trained in the PPMI sample and then tested in the clinical sample is summarized in Table 3. Cross-validated performance of HV-SBR with cut-off optimized in the clinical sample is given as benchmark for comparison.

Classification performance of the semi-quantitative SBR measures and the CNN in the different settings of the clinical sample is summarized in Table 4. The accuracy of the SBR measures did not differ between filtered backprojection and iterative reconstruction (AAL-SBR: $p = 0.125$; HV-SBR: $p = 0.563$) but was lower in the mixed setting. The reduction of AAL-SBR accuracy in the mixed setting was statistically significant compared with the filtered backprojection setting ($p = 0.004$). It did not reach statistical significance compared with the iterative reconstruction setting ($p = 0.188$). The reduction of HV-SBR accuracy in the mixed clinical setting was statistically significant compared with both pure settings (filtered backprojection $p = 0.016$, iterative reconstruction $p = 0.043$). By contrast, CNN accuracy did not differ between the mixed setting and any of the pure settings. There was even a tendency towards higher CNN accuracy in the mixed clinical setting compared with the iterative reconstruction setting ($p = 0.094$).

Discussion

The major finding of this study is that a high-dimensional CNN can be trained to deal with variable image characteristics in FP-CIT SPECT considerably better than conventional SBR analysis. This was first shown in the mixed PPMI setting comprising original and heavily pre-smoothed SPECT images. Overall accuracy dropped for both SBR measures (from 0.957 ± 0.017 in the original PPMI setting to 0.900 ± 0.029 in

Table 1 Diagnostic performance of the different putaminal SBR measures and the CNN in the original (unsmoothed) PPMI setting. For each split of the PPMI data, SBR cut-offs and CNN parameters were optimized in the training sample, the performance measures were estimated by applying these models to the test sample. Given are mean ± standard deviation over the random splits

Method	PPMI-SBR	AAL-SBR	HV-SBR	CNN
Cut-off	1.098 ± 0.058	1.007 ± 0.044	0.782 ± 0.066	
Accuracy	0.970 ± 0.017	0.957 ± 0.017	0.966 ± 0.011	0.972 ± 0.014
Sensitivity	0.954 ± 0.027	0.927 ± 0.033	0.946 ± 0.019	0.983 ± 0.012
Specificity	0.985 ± 0.021	0.988 ± 0.017	0.985 ± 0.026	0.962 ± 0.024

Table 2 Diagnostic performance of conventional putamen SBR (AAL-SBR), hottest voxels putamen SBR (HV-SBR), and the convolutional neural network (CNN) in original (unsmoothed) PPMI setting, 18-mm pre-smoothed PPMI setting, and mixed PPMI setting comprising original and pre-smoothed images. For each split of the PPMI data, SBR cut-offs and CNN parameters were fixed in the training sample, the performance measures were estimated by applying these models to the test sample. Given are mean ± standard deviation over the random splits

Method	AAL-SBR			HV-SBR			CNN		
Sample	Original	Smoothed	Mixed	Original	Smoothed	Mixed	Original	Smoothed	Mixed
Cut-off	1.007 ± 0.044	0.664 ± 0.024	0.743 ± 0.044	0.782 ± 0.066	0.584 ± 0.027	0.646 ± 0.032			
Accuracy	0.957 ± 0.017**	0.952 ± 0.015***	0.900 ± 0.029	0.966 ± 0.011*	0.957 ± 0.015	0.951 ± 0.010	0.972 ± 0.014	0.955 ± 0.009	0.967 ± 0.018
Sensitivity	0.927 ± 0.033	0.924 ± 0.027	0.863 ± 0.058	0.946 ± 0.019	0.937 ± 0.029	0.927 ± 0.016	0.983 ± 0.012	0.934 ± 0.012	0.966 ± 0.024
Specificity	0.988 ± 0.017	0.980 ± 0.015	0.937 ± 0.046	0.985 ± 0.026	0.978 ± 0.021	0.976 ± 0.026	0.962 ± 0.024	0.976 ± 0.024	0.968 ± 0.033

* Different compared with the accuracy of the same classification method in the mixed setting: two-sided Wilcoxon signed rank test $p < 0.05$

** Different compared with the accuracy of the same classification method in the mixed setting: two-sided Wilcoxon signed rank test $p < 0.01$

the mixed PPMI setting for the conventional putamen AAL-SBR, from 0.966 ± 0.011 to 0.951 ± 0.010 for HV-SBR, Table 2). By contrast, the CNN showed about the same accuracy in the mixed PPMI setting (0.967 ± 0.018) as in both pure settings (0.972 ± 0.014 and 0.955 ± 0.009, Table 2). These findings were confirmed in the clinical data comprising SPECT images reconstructed with two different reconstruction algorithms, filtered backprojection, and iterative ordered-subsets-expectation-maximization with resolution recovery. CNN accuracy did not differ between the mixed clinical setting and any of the pure clinical settings, whereas both SBR measures performed worse in the mixed setting compared with the pure clinical settings (Table 4).

The robustness of the CNN to variable image characteristics was also demonstrated by applying models trained in the mixed PPMI setting to the completely independent clinical patient sample. The PPMI-trained HV-SBR did not show acceptable performance in the clinical sample whereas the PPMI-trained CNN did (Table 3). The fact that the performance of the PPMI-trained CNN in the clinical sample was very similar for filtered backprojection and iterative reconstruction of the clinical sample further supports the robustness of the CNN with respect to variable image characteristics, because the backprojected images (without scatter correction) and the iteratively reconstructed images (with scatter correction and resolution recovery) were quite different not only with respect to uniformity of FP-CIT concentration in extrastriatal brain regions but also with respect to striatum-to-background contrast (Fig. 2). The latter resulted in a considerably higher cut-off for optimal differentiation between neurodegenerative and non-neurodegenerative PS in the clinical sample by HV-SBR in case of iterative reconstruction compared with filtered backprojection (1.232 ± 0.088 versus 0.882 ± 0.012, Table 3).

The good CNN performance in the mixed settings (PPMI and clinical) and of the PPMI-trained CNN in the clinical settings most likely is associated with the use of convolutional layers, at least to some extent. One or several of the convolutional layers may have learned to factor out non-diagnostic image characteristics, like resolution-related smoothing effects, thereby allowing later layers in the network feature extraction from the same image baseline for all images in the training set. This property of CNNs to separate so-called latent nuisance parameters from the confounding factors of the learning objective is well known, and as of late has been theoretically linked to Bayesian networks [47]. Thus, CNN can be assumed to be particularly appropriate to deal with site- and camera-specific variability of spatial resolution in FP-CIT SPECT, as soon as they exhibit sufficient depth and a sufficient number of parameters represented in filter kernel weights.

A custom CNN architecture was used in the present study (Fig. 4). Transfer learning from an Inception V3 network [48]

Table 3 Diagnostic performance of the hottest voxels putamen SBR (HV-SBR) with cut-offs optimized in different PPMI training samples (Table 2) and of the CNN trained in the mixed PPMI sample when applied to the clinical sample with filtered backprojection (FBP, A) or iterative reconstruction with resolution recovery (IR, B). Cross-validated performance of the putamen HV-SBR using a cut-off optimized in the clinical sample is given as benchmark (Table 4)

	Method Training sample	HV-SBR Clinical (FBP)	HV-SBR Original PPMI	HV-SBR Smoothed PPMI	HV-SBR Mixed PPMI	CNN Mixed PPMI
A	Cut-off	0.882 ± 0.012	0.782	0.584	0.646	
	Accuracy	0.980 ± 0.015	0.950	0.815	0.866	0.976
	Sensitivity	0.963 ± 0.025	0.906	0.631	0.732	0.959
	Specificity	0.997 ± 0.011	0.993	1.000	1.000	0.993
B	Cut-off	1.232 ± 0.088	0.782	0.584	0.646	
	Accuracy	0.982 ± 0.017	0.862	0.695	0.752	0.970
	Sensitivity	0.967 ± 0.031	0.725	0.389	0.503	0.939
	Specificity	0.997 ± 0.011	1.000	1.000	1.000	1.000

pre-trained on ImageNet [49] was also explored (Supplementary "Transfer learning from an Inception V3 network for classification of FP-CIT SPECT"). The validation loss systematically increased when training loss converged, strongly indicating overfitting on the training data. This lets us hypothesize that the features learned by InceptionV3 on natural images are not optimal for classification of FP-CIT images of much lower resolution and size, and that the number of features Inception V3 generates is too high compared to the number of training cases used for training or finetuning of the network. In order to support further testing of transfer learning from an Inception V3 network, we share our program code and the data via GitHub (https://github.com/mtwenzel/parkinson-classification). The code is provided in the free Colaboratory Jypyter notebook environment to allow running and modifying the code online (https://colab.research.google.com/github/mtwenzel/parkinson-classification/blob/master/PPMI-InceptionV3.ipynb).

It should be mentioned that conventional machine learning methods (e.g., logistic regression or support vector machines) may be able to perform on an equal level as CNN, assuming a controllable and a priori known degree of variation in image characteristics, and assuming a profound human-engineered set of image features. However, the point in using a feature learning approach instead of a feature engineering approach was to open a path towards automatic extraction of the important features not only across pathological variability but also across scanners and reconstruction algorithms.

CNN performance in the pure settings was slightly better compared to the SBR measures in the PPMI data (Table 2) and slightly worse in the clinical data (Table 4). This might be explained by the smaller size of the clinical sample compared with the PPMI sample.

A secondary finding of this study was that the performance of the CNN trained in the smoothed PPMI setting showed acceptable performance in the original PPMI setting (accuracy = 0.945 ± 0.058) whereas the CNN trained in the original PPMI data performed badly in the smoothed PPMI setting (accuracy = 0.629 ± 0.069). This suggests that a CNN trained in FP-CIT images with a low spatial resolution is useful also for classification of SPECT images with higher spatial resolution, but not vice versa. This is a partial improvement from SBR measures which performed badly in both PPMI cross-site settings (results not shown).

Only a few previous studies tested deep learning for classification of FP-CIT SPECT [50–52]. Choi and co-workers developed a deep learning-based system for classification of 3-dimensional FP-CIT SPECT images, called "PD Net" [50]. Cross-validated accuracy of PD Net in PPMI data was 0.960, in good agreement with the present results (Table 1). Good performance of PD Net was confirmed in an independent clinical dataset acquired at the authors' site. The study did not assess the impact of spatial resolution on CNN performance. Kim and co-workers provided proof of concept for the use of transfer learning, i.e., limited retraining of CNNs pre-trained on nonmedical images for classification of FP-CIT SPECT [51].

The problem of variable image characteristics depending on acquisition hardware and/or reconstruction software used for image generation is not specific to FP-CIT SPECT but affects many types of medical images. For example, MRI-derived brain volumetric measures are promising markers to support the diagnosis of various neurological and psychiatric diseases [53]. However, MRI-derived brain volumetric measures are affected by scanner-related factors, including scanner field strength, manufacturer, and details of the acquisition sequence. For example, cross-field strength (1.5 T versus 3 T) intra-subject variability of MRI-derived global cortical thickness is about 6 times larger than within-scanner variability when the subject is scanned twice on the same scanner [54]. This limits the use of MRI-derived brain volumetric measures in clinical routine, very similar to conventional semi-

Table 4 Diagnostic performance of the conventional putamen SBR (AAL-SBR), the hottest voxels putamen SBR (HV-SBR), and the convolutional neural network (CNN) in the clinical sample with filtered backprojection, iterative reconstruction, or mixed reconstruction (filtered backprojection or iterative reconstruction selected randomly). For each split of the clinical data, the SBR cut-offs and CNN parameters were optimized in the training (and validation) sample, the performance measures were estimated by applying these models to the test sample. Given are mean ± standard deviation over the random splits

Method	AAL-SBR			HV-SBR			CNN		
Reconstruction	Filtered backprojection	Iterative	Mixed	Filtered backprojection	Iterative	Mixed	Filtered backprojection	Iterative	Mixed
Cut-off	1.191 ± 0.030	1.594 ± 0.082	1.361 ± 0.041	0.882 ± 0.012	1.232 ± 0.088	1.009 ± 0.059			
Accuracy	0.983 ± 0.018**	0.967 ± 0.025	0.950 ± 0.016	0.980 ± 0.015*	0.982 ± 0.017*	0.960 ± 0.020	0.960 ± 0.015	0.960 ± 0.027	0.977 ± 0.016
Sensitivity	0.977 ± 0.027	0.947 ± 0.042	0.923 ± 0.039	0.963 ± 0.025	0.967 ± 0.031	0.927 ± 0.038	0.946 ± 0.053	0.940 ± 0.049	0.960 ± 0.026
Specificity	0.990 ± 0.016	0.987 ± 0.028	0.977 ± 0.027	0.997 ± 0.011	0.997 ± 0.011	0.993 ± 0.014	0.975 ± 0.030	0.980 ± 0.032	0.993 ± 0.014

* Different compared to the accuracy of the same classification method in the mixed setting: two-sided Wilcoxon signed rank test $p < 0.05$

** Different compared to the accuracy of the same classification method in the mixed setting: two-sided Wilcoxon signed rank test $p < 0.01$

quantitative measures derived from FP-CIT SPECT. Several groups have shown that deep neural networks are very promising to address this problem ("domain adaption") in brain MRI [55–58].

Limitations of the present study include the following. First, the study did not address all aspects of image characteristics but focused on spatial resolution (e.g., the impact of different attenuation correction methods was not assessed). However, spatial resolution is the major factor contributing to site- and camera-specific variability of image characteristics in nuclear medicine brain imaging [59], suggesting that the mixed PPMI setting including original (unsmoothed) images and heavily smoothed is not too unrealistic. This was supported by the good performance of the CNN trained in the mixed PPMI setting in the different settings of the clinical patient sample. We consider these findings as proof of concept that CNNs can be trained to be robust with respect to site-, camera-, and scan-specific variability of image characteristics in FP-CIT SPECT. This was confirmed in a fully independent patient sample from clinical routine patient care. We do not expect the CNN trained in the mixed PPMI sample to be ready for use in a large variety of different settings. Second, our CNN used 2-dimensional slab views to reduce computation time for training whereas SBR analysis used the 3-dimensional SPECT images. CNN performance might further improve when using 3-dimensional images. Third, the CNN provided a score ranging between 0 and 1 for each SPECT image to belong to a given class. For the analyses presented here, categorization was based on a cut-off of 0.5. The actual value of the score was not taken into account, although it might be useful information whether the actual value is 0.60 or 0.99. Furthermore, CNN performance might be optimized for a given task by selecting a task-specific cut-off for classification. For example, fully automatic CNN-based classification may serve as primary (rather than second or third) reader to identify clearly normal FP-CIT SPECT images that require no further interpretation. This could reduce the workload of the clinician. Highest possible sensitivity even at the cost of reduced specificity is required for this task.

In conclusion, this study provides proof of concept that a high-dimensional CNN can be trained to be robust with respect to variable site-, camera-, or scan-specific image characteristics in FP-CIT SPECT without a large loss of diagnostic accuracy compared to mono-site/mono-camera settings, in contrast to the conventional SBR. Based on this, we hypothesize that a single CNN can be used to support the interpretation of FP-CIT SPECT at many different sites using different acquisition hardware and/or reconstruction software with only minor harmonization of acquisition and reconstruction protocols. We propose a common effort of the nuclear medicine community to retrospectively collect FP-CIT SPECT data covering a wide range of acquisition and reconstruction protocols to test this hypothesis.

Supplementary: transfer learning from an inception V3 network for classification of FP-CIT SPECT

We also explored transfer learning from an Inception V3 network [48] pre-trained on ImageNet [49] for classification of FP-CIT SPECT images. A number of best practice settings with and without data augmentation were used. We tested several numbers of added dense layers after the pre-trained inception blocks, settling with a number that yields a number of trainable parameters slightly lower than the number of free parameters in our proposed CNN architecture (about 2–4 Mio. depending on the actual experiment). The rationale was that most of the free parameters in our proposed CNN will be tuned for feature extraction rather than classification from the features; hence, a low number of trainable parameters in the added dense layers after the inception blocks will prevent overfitting. Further, we explored two ways of training: either only training the added dense layers, or first training the dense layers with a higher learning rate, then training further the last one or two inception blocks plus the dense layers with a lowered training rate. In all experiments, the validation loss systematically increased when training loss converged, strongly indicating overfitting on the training data. This lets us hypothesize that the features learned by InceptionV3 on natural images are not optimal for classification of FP-CIT images of much lower resolution and size, and that the number of features Inception V3 generates is too high compared to the number of training cases used for training or fine tuning of the network. This is in line with the fact that transfer learning from real-world color images to medical grayscale images today is widely considered sub-optimal.

Kim and co-workers also explored transfer learning of an Inception V3 network for automatic classification of FP-CIT SPECT [51]. The overall accuracy they achieved was 0.844 (computed from Table 1 in [51]). This accuracy is not very high, given the fact that the standard of truth for the data used in this study was a visual interpretation of the FP-CIT SPECT images (rather than clinical diagnosis) and that visually inconclusive SPECT images were excluded. It suggests that the transfer learning of an InceptionV3 network for classification of FP-CIT SPECT performed in this study also did not result in a network that provides acceptable performance to be used in clinical patient care.

In summary, we believe that the potential of transfer learning of powerful large networks pre-trained for classification of color photos compared to full training of smaller networks (as in the present study) is still an open question in FP-CIT SPECT that requires further investigation. In order to support this, we share the code and the data of our Inception V3 transfer learning tests via GitHub (https://github.com/mtwenzel/parkinson-classification). The code is provided in the free Colaboratory Jupyter notebook environment to allow running and modifying the code online (https://colab.research.google.com/github/mtwenzel/parkinson-classification/blob/master/PPMI-InceptionV3.ipynb).

Acknowledgments PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners including the following: Abbvie, Avid Radiopharmaceuticals, Biogen, BioLegend, Bristol-Myers Squibb, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, Sanofi Genzyme, Servier, Takeda, Teva, and UCB. For up-to-date information about all of the PPMI funding partners, visit www.ppmi-info.org/fundingpartners.

Compliance with ethical standards

Ethical approval Waiver of informed consent for the retrospective analyses of the anonymized clinical data in this study was obtained from the ethics review board of the general medical council of the state of Hamburg, Germany. All procedures performed in this study were in accordance with the ethical standards of the ethics review board of the general medical council of the state of Hamburg, Germany, and with the 1964 Helsinki declaration and its later amendments.

Conflict of interest F.M. is employee of Nvidia, J. K is employee of Jung Diagnostics, and M.E. is employee of Pinax Pharma. This did not influence the content of this manuscript, neither directly nor indirectly. There is no actual or potential conflict of interest for any of the other authors.

References

1. Booij J, Speelman JD, Horstink MW, Wolters EC. The clinical benefit of imaging striatal dopamine transporters with [123I]FP-CIT SPET in differentiating patients with presynaptic parkinsonism from those with other forms of parkinsonism. Eur J Nucl Med. 2001;28:266–72.
2. Darcourt J, Booij J, Tatsch K, Varrone A, Vander Borght T, Kapucu OL, et al. EANM procedure guidelines for brain neurotransmission SPECT using (123)I-labelled dopamine transporter ligands, version 2. Eur J Nucl Med Mol Imaging. 2010;37:443–50. https://doi.org/10.1007/s00259-009-1267-x.
3. Tatsch K, Poepperl G. Nigrostriatal dopamine terminal imaging with dopamine transporter SPECT: an update. J Nucl Med. 2013;54:1331–8. https://doi.org/10.2967/jnumed.112.105379.
4. Van Laere K, Everaert L, Annemans L, Gonce M, Vandenberghe W, Vander Borght T. The cost effectiveness of 123I-FP-CIT SPECT imaging in patients with an uncertain clinical diagnosis of parkinsonism. Eur J Nucl Med Mol Imaging. 2008;35:1367–76. https://doi.org/10.1007/s00259-008-0777-2.
5. O'Brien JT, Oertel WH, McKeith IG, Grosset DG, Walker Z, Tatsch K, et al. Is ioflupane I123 injection diagnostically effective in patients with movement disorders and dementia? Pooled analysis of four clinical trials. BMJ Open. 2014;4:ARTN e005122. https://doi.org/10.1136/bmjopen-2014-005122.
6. Seibyl JP, Kupsch A, Booij J, Grosset DG, Costa DC, Hauser RA, et al. Individual-reader diagnostic performance and between-reader agreement in assessment of subjects with Parkinsonian syndrome or dementia using 123I-ioflupane injection (DaTscan) imaging. J Nucl Med. 2014;55:1288–96. https://doi.org/10.2967/jnumed.114.140228.
7. Booij J, Dubroff J, Pryma D, Yu J, Agarwal R, Lakhani P, et al. Diagnostic performance of the visual reading of I-123-ioflupane

SPECT images with or without quantification in patients with movement disorders or dementia. J Nucl Med. 2017;58:1821–6. https://doi.org/10.2967/jnumed.116.189266.
8. Soderlund TA, Dickson JC, Prvulovich E, Ben-Haim S, Kemp P, Booij J, et al. Value of semiquantitative analysis for clinical reporting of 123I-2-beta-carbomethoxy-3beta-(4-iodophenyl)-N-(3-fluoropropyl)nortropane SPECT studies. J Nucl Med. 2013;54:714–22. https://doi.org/10.2967/jnumed.112.110106.
9. Badiavas K, Molyvda E, Iakovou I, Tsolaki M, Psarrakos K, Karatzas N. SPECT imaging evaluation in movement disorders: far beyond visual assessment. Eur J Nucl Med Mol Imaging. 2011;38:764–73. https://doi.org/10.1007/s00259-010-1664-1.
10. Tatsch K, Poepperl G. Quantitative approaches to dopaminergic brain imaging. Q J Nucl Med Mol Imaging. 2012;56:27–38.
11. Oliveira FPM, Faria DB, Costa DC, Castelo-Branco M, Tavares J. Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of Parkinson's disease based on [(123)I]FP-CIT SPECT images. Eur J Nucl Med Mol Imaging. 2018;45:1052–62. https://doi.org/10.1007/s00259-017-3918-7.
12. Nobili F, Naseri M, De Carli F, Asenbaum S, Booij J, Darcourt J, et al. Automatic semi-quantification of [123I]FP-CIT SPECT scans in healthy volunteers using BasGan version 2: results from the ENC-DAT database. Eur J Nucl Med Mol Imaging. 2013;40:565–73. https://doi.org/10.1007/s00259-012-2304-8.
13. Albert NL, Unterrainer M, Diemling M, Xiong GM, Bartenstein P, Koch W, et al. Implementation of the European multicentre database of healthy controls for [I-123]FP-CIT SPECT increases diagnostic accuracy in patients with clinically uncertain parkinsonian syndromes. Eur J Nucl Med Mol Imaging. 2016;43:1315–22. https://doi.org/10.1007/s00259-015-3304-2.
14. Dickson JC, Tossici-Bolt L, Sera T, Booij J, Ziebell M, Morbelli S, et al. The impact of reconstruction and scanner characterisation on the diagnostic capability of a normal database for [123I]FP-CIT SPECT imaging. EJNMMI Res. 2017;7:10. https://doi.org/10.1186/s13550-016-0253-0.
15. Fujita M, Varrone A, Kim KM, Watabe H, Zoghbi SS, Baldwin RM, et al. Effect of scatter correction in the measurement of striatal and extrastriatal dopamine D2 receptors using (123)Iepidepride SPECT. J Nucl Med. 2001;42:217.
16. Lange C, Seese A, Schwarzenbock S, Steinhoff K, Umland-Seidler B, Krause BJ, et al. CT-based attenuation correction in I-123-ioflupane SPECT. PLoS One. 2014;9:e108328. https://doi.org/10.1371/journal.pone.0108328.
17. Meyer PT, Sattler B, Lincke T, Seese A, Sabri O. Investigating dopaminergic neurotransmission with I-123-FP-CIT SPECT: comparability of modern SPECT systems. J Nucl Med. 2003;44:839–45.
18. Tossici-Bolt L, Dickson JC, Sera T, Booij J, Asenbaun-Nan S, Bagnara MC, et al. [123I]FP-CIT ENC-DAT normal database: the impact of the reconstruction and quantification methods. EJNMMI Phys. 2017;4:8. https://doi.org/10.1186/s40658-017-0175-6.
19. Tossici-Bolt L, Dickson JC, Sera T, de Nijs R, Bagnara MC, Jonsson C, et al. Calibration of gamma camera systems for a multicentre European (1)(2)(3)I-FP-CIT SPECT normal database. Eur J Nucl Med Mol Imaging. 2011;38:1529–40. https://doi.org/10.1007/s00259-011-1801-5.
20. Varrone A, Dickson JC, Tossici-Bolt L, Sera T, Asenbaum S, Booij J, et al. European multicentre database of healthy controls for [123I]FP-CIT SPECT (ENC-DAT): age-related effects, gender differences and evaluation of different methods of analysis. Eur J Nucl Med Mol Imaging. 2013;40:213–27. https://doi.org/10.1007/s00259-012-2276-8.
21. Buchert R, Kluge A, Tossici-Bolt L, Dickson J, Bronzel M, Lange C, et al. Reduction in camera-specific variability in [(123)I]FP-CIT SPECT outcome measures by image reconstruction optimized for multisite settings: impact on age-dependence of the specific binding ratio in the ENC-DAT database of healthy controls. Eur J Nucl Med Mol Imaging. 2016;43:1323–36. https://doi.org/10.1007/s00259-016-3309-5.
22. Koch W, Bartenstein P, la Fougere C. Radius dependence of FP-CIT quantification: a Monte Carlo-based simulation study. Ann Nucl Med. 2014;28:103–11. https://doi.org/10.1007/s12149-013-0789-2.
23. Koch W, Mustafa M, Zach C, Tatsch K. Influence of movement on FP-CIT SPECT quantification: a Monte Carlo based simulation. Nucl Med Commun. 2007;28:603–14. https://doi.org/10.1097/MNM.0b013e328273bc6f.
24. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44. https://doi.org/10.1038/nature14539.
25. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88. https://doi.org/10.1016/j.media.2017.07.005.
26. Choi H, Jin KH. Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. Behav Brain Res. 2018;344:103–9. https://doi.org/10.1016/j.bbr.2018.02.017.
27. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, et al. A deep learning model to predict a diagnosis of Alzheimer disease by using (18)F-FDG PET of the brain. Radiology. 2018:180958. https://doi.org/10.1148/radiol.2018180958.
28. Liu M, Cheng D, Yan W. Alzheimer's disease neuroimaging I. classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. Front Neuroinform. 2018;12:35. https://doi.org/10.3389/fninf.2018.00035.
29. Lu D, Popuri K, Ding GW, Balachandar R, Beg MF. Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. Med Image Anal. 2018;46:26–34. https://doi.org/10.1016/j.media.2018.02.002.
30. Segovia F, Gorriz JM, Ramirez J, Martinez-Murcia FJ, Garcia-Perez M. Using deep neural networks along with dimensionality reduction techniques to assist the diagnosis of neurodegenerative disorders. Log J IGPL. 2018;26:618–28. https://doi.org/10.1093/jigpal/jzy026.
31. Shi J, Zheng X, Li Y, Zhang Q, Ying S. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. IEEE J Biomed Health Inform. 2018;22:173–83. https://doi.org/10.1109/jbhi.2017.2655720.
32. Zhou T, Thung KH, Zhu X, Shen D. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Hum Brain Mapp. 2018. https://doi.org/10.1002/hbm.24428.
33. Hirata K, Takeuchi W, Yamaguchi S, Kobayashi H, Terasaka S, Toyonaga T, et al. Convolutional neural network can help differentiate FDG PET images of brain tumor between glioblastoma and primary central nervous system lymphoma. J Nucl Med. 2016;57.
34. Blanc-Durand P, Van der Gucht A, Schaefer N, Itti E, Prior JO. Automatic lesion detection and segmentation of F-18-FET PET in gliomas: a full 3D U-net convolutional neural network study. PLoS One. 2018;13:ARTN e0195798. https://doi.org/10.1371/journal.pone.0195798.
35. Kang SK, Seo S, Shin SA, Byun MS, Lee DY, Kim YK, et al. Adaptive template generation for amyloid PET using a deep learning approach. Hum Brain Mapp. 2018;39:3769–78. https://doi.org/10.1002/hbm.24210.
36. Xiang L, Qiao Y, Nie D, An L, Lin WL, Wang Q, et al. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing. 2017;267:406–16. https://doi.org/10.1016/j.neucom.2017.06.048.

37. Wang Y, Yu BT, Wang L, Zu C, Lalush DS, Lin WL, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. Neuroimage. 2018;174:550–62. https://doi.org/10.1016/j.neuroimage.2018.03.045.
38. Li RJ, Zhang WL, Suk HI, Wang L, Li J, Shen DG, et al. Deep learning based imaging data completion for improved brain disease diagnosis. Lect Notes Comput Sc. 2014;8675:305–12.
39. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep learning MR imaging-based attenuation correction for PET/MR imaging. Radiology. 2018;286:676–84. https://doi.org/10.1148/radiol.2017170700.
40. Parkinson Progression Marker I. The Parkinson progression marker initiative (PPMI). Prog Neurobiol. 2011;95:629–35. https://doi.org/10.1016/j.pneurobio.2011.09.005.
41. 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. https://doi.org/10.2967/jnumed.111.100784.
42. Yokoyama K, Imabayashi E, Sumida K, Sone D, Kimura Y, Sato N, et al. Computed-tomography-guided anatomic standardization for quantitative assessment of dopamine transporter SPECT. Eur J Nucl Med Mol Imaging. 2017;44:366–72. https://doi.org/10.1007/s00259-016-3496-0.
43. Kupitz D, Apostolova I, Lange C, Ulrich G, Amthauer H, Brenner W, et al. Global scaling for semi-quantitative analysis in FP-CIT SPECT. Nuklearmed-Nucl Med. 2014;53:234–41. https://doi.org/10.3413/Nukmed-0659-14-04.
44. Koch W, Unterrainer M, Xiong G, Bartenstein P, Diemling M, Varrone A, et al. Extrastriatal binding of [(1)(2)(3)I]FP-CIT in the thalamus and pons: gender and age dependencies assessed in a European multicentre database of healthy controls. Eur J Nucl Med Mol Imaging. 2014;41:1938–46. https://doi.org/10.1007/s00259-014-2785-8.
45. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 2002;15:273–89. https://doi.org/10.1006/nimg.2001.0978.
46. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3:32–5.
47. Patel AB, Nguyen T, Baraniuk RG. A probabilistic framework for deep learning. Adv Neur In. 2016;29.
48. Szegedy C, Vanhouke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: IEEE conference on computer vision and pattern recognition; 2015. p. 2818–26.
49. Russakovsky O, Deng J, Su H, Krause J, Sateesh S, Ma S, et al. ImageNet large scale visual recognition challenge. J Comput Vision. 2014;115:1–42.
50. Choi H, Ha S, Im HJ, Paek SH, Lee DS. Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. Neuroimage Clin. 2017;16:586–94. https://doi.org/10.1016/j.nicl.2017.09.010.
51. Kim DH, Wit H, Thurston M. Artificial intelligence in the diagnosis of Parkinson's disease from ioflupane-123 single-photon emission computed tomography dopamine transporter scans using transfer learning. Nucl Med Commun. 2018;39:887–93. https://doi.org/10.1097/MNM.0000000000000890.
52. Martinez-Murcia FJ, Gorriz JM, Ramirez J, Ortiz A. Convolutional neural networks for neuroimaging in Parkinson's disease: is preprocessing needed? Int J Neural Syst. 2018:1850035. https://doi.org/10.1142/S0129065718500351.
53. Giorgio A, De Stefano N. Clinical use of brain volumetry. J Magn Reson Imaging. 2013;37:1. https://doi.org/10.1002/jmri.23671.
54. Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage. 2006;32:180–94. https://doi.org/10.1016/j.neuroimage.2006.02.051.
55. Kouw WM, Loog M, Bartels LW, Mendrik AM. MR acquisition invariant representation learning. arXiv. 2018;arXiv:1709.07944v2.
56. Ghafoorian M, Mehrtash A, Kapur T, Karssemeijer N, Marchiori E, Pesteie M, et al. Transfer learning for domain adaptation in MRI: application in brain lesion segmentation. arXiv. 2017; https://arxiv.org/pdf/1702.07841.pdf.
57. Hosseini-Asl E, Keynton R, El-Baz A. Alzheimer's disease diagnostics by adaption of 3D convolutional network. arXiv. 2016; https://arxiv.org/pdf/1702.07841.pdf.
58. Valverde S, Salem M, Cabezas M, Pareto D, Vilanova JC, Ramio-Torrenta L, et al. One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks. Neuroimage Clin. 2019;21:101638. https://doi.org/10.1016/j.nicl.2018.101638.
59. Joshi A, Koeppe RA, Fessler JA. Reducing between scanner differences in multi-center PET studies. Neuroimage. 2009;46:154–9. https://doi.org/10.1016/j.neuroimage.2009.01.057.