From Sinograms to Diagnosis: Task-Aware Evaluation of Reconstruction and Direct Classification in Parkinsonian DaT-SPECT

Anonymized Authors[1]

[1] Anonymized Affiliations
email@anonymized.com

Abstract. Dopamine transporter single-photon emission computed tomography (DaT-SPECT) is typically interpreted after image reconstruction, and most artificial intelligence pipelines follow the same image-domain workflow. However, the representation best suited for physician interpretation may not be the representation most effective for automated diagnosis. We compared five diagnostic pathways: classification of original images, direct U-Net reconstruction from sinograms, ordered-subsets expectation maximization (OSEM) reconstruction, OSEM followed by Residual U-Net refinement, and direct classification of image-derived sinograms. Experiments were conducted on two PPMI-derived two-dimensional DaT-SPECT datasets across PD versus non-PD, PD versus control, and PD versus SWEDD tasks. Reconstruction quality was assessed using PSNR and global SSIM, while diagnostic performance was evaluated using ResNet50, VGG16, Inception-v3, and Xception. OSEM followed by residual refinement achieved the highest reconstruction quality and generally improved classification performance relative to OSEM alone on the larger dataset. Direct sinogram classification matched the highest observed accuracy of the refined reconstruction pathway across all evaluated tasks. Nevertheless, an exploratory qualitative review by two physicians suggested that some directly reconstructed images contained local uptake alterations that were not fully reflected by global image-quality metrics or downstream classification performance. These findings indicate that reconstruction quality, automated diagnostic performance, and suitability for physician interpretation are related but not interchangeable. OSEM-based residual refinement provides the strongest image-domain representation, whereas direct sinogram analysis may serve as a complementary pathway for automated DaT-SPECT classification.

Keywords: Parkinson's disease, SPECT reconstruction, Medical imaging.

1 Introduction

Dopamine transporter single-photon emission computed tomography (DaT-SPECT) provides functional information about presynaptic dopaminergic activity and is widely used to support the assessment of Parkinsonian disorders [1, 2]. Reduced striatal uptake

is associated with nigrostriatal dopaminergic degeneration, while scans without evidence of dopaminergic deficit may indicate an alternative cause of Parkinsonian symptoms.

SPECT measurements are acquired in the projection domain and reconstructed into spatial images before clinical interpretation. Ordered-subsets expectation maximization (OSEM) is commonly used because it provides an efficient iterative approximation of maximum-likelihood reconstruction [3]. However, reconstruction modifies noise, contrast, and local uptake patterns. These changes may improve visual interpretation while also altering information available to a downstream classifier.

Deep learning has been investigated for both projection-to-image reconstruction and DaT-SPECT classification. AI reconstruction methods are usually evaluated using image-quality metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [4–7]. Classification studies, in contrast, usually begin with reconstructed images and report diagnostic metrics such as accuracy, AUROC, and F1-score [8–11]. The relationship between these two objectives remains insufficiently studied. A reconstruction with higher image similarity may not necessarily provide a better diagnostic representation, while a non-image representation may still contain information useful for automated classification.

Previous study evaluated the effect of training protocols on CNN-based DaT-SPECT classification [12]. The present work extends this framework by comparing image- and sinogram-domain representations within the same diagnostic tasks.

Five pathways are evaluated:

1. Original-image classification.
2. Direct sinogram-to-image reconstruction followed by classification.
3. OSEM reconstruction followed by classification.
4. OSEM followed by Residual U-Net refinement and classification.
5. Direct sinogram classification without image reconstruction.

The main objective is to determine whether the representation producing the most accurate reconstructed image is also the representation that provides the strongest downstream classification.

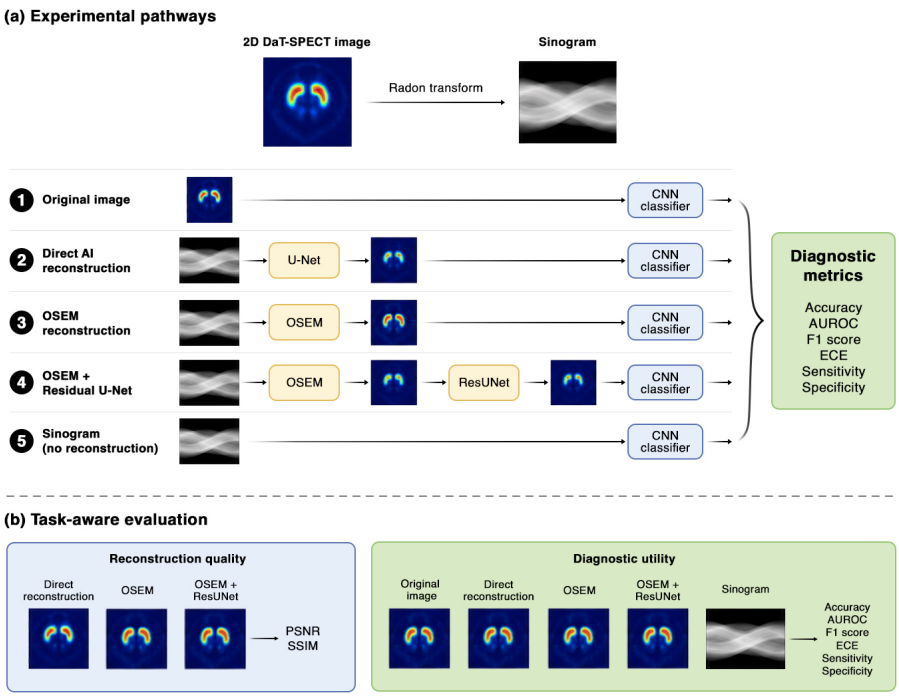

Fig. 1. Task-aware DaT-SPECT framework. (a) Five pathways were evaluated using original images and Radon-derived sinograms. (b) Reconstructed images were assessed with PSNR and global SSIM, while all pathways were evaluated by downstream classification.

2 Materials and Methods

2.1 Datasets and Diagnostic Tasks

Two previously curated PPMI-derived two-dimensional DaT-SPECT datasets were used. The first, denoted PPMI-PD Small, contained 645 images and was used for PD versus non-PD classification. The second, denoted PPMI-PD Large, contained 1,528 images belonging to the PD, Control, and SWEDD categories.

The released two-dimensional images were used directly, without repeating the original volume-to-image extraction procedures. The PPMI-PD Small dataset followed the preprocessing and partitioning described by Kurmi et al. [8], whereas the PPMI-PD Large images and partitions were prepared according to the protocol reported in a separate study [12].

Three binary tasks were constructed from PPMI-PD Large: PD versus non-PD, PD versus control, and PD versus SWEDD. SWEDD refers to scans without evidence of dopaminergic deficit. This comparison is clinically relevant because subjects may present symptoms suggestive of Parkinsonism despite the absence of the characteristic dopaminergic imaging pattern.

Table 1. Dataset and task summary. Dimensions are reported as height × width.

Setting	Image size	Sinogram size	Task	Train/Test
PPMI-PD Small	224×224	224×180	PD vs. non-PD	PD 346/86; non-PD 170/43
BigNPD	733×612	612×180	PD vs. non-PD	PD 928/225; non-PD 304/71
BigSWEDD	733×612	612×180	PD vs. SWEDD	PD 928/225; SWEDD 64/15
BigCONTROL	733×612	612×180	PD vs. control	PD 928/225; control 240/56

Every derived representation retained the partition assignment of its source image. Therefore, an original image and its corresponding sinogram and reconstructed versions were always assigned to the same experimental split.

2.2 Sinogram Generation

Each released image was intensity-normalized and transformed using the discrete Radon transform with 180 uniformly distributed projection angles. This produced sinograms of size 224×180 for PPMI-PD Small and 612×180 for PPMI-PD Large.

The resulting sinograms were generated from already reconstructed images rather than obtained directly from a scanner. They should therefore be interpreted as controlled image-derived Radon representations. They do not reproduce the complete clinical acquisition process, including attenuation, scatter, collimator response, motion, or acquisition-level photon noise.

2.3 Reconstruction Pathways

A modified U-Net encoder-decoder was trained for direct sinogram-to-image reconstruction [13]. Separate models were used for the two image resolutions. The Small model mapped 224×180 sinograms to 224×224 images, whereas the Large model mapped 612×180 sinograms to 733×612 images. The models were trained for 200 epochs using mean squared error and Adam optimization with a learning rate of 10^{-4}.

OSEM was applied to the same sinograms as the iterative reconstruction baseline. A fixed configuration of 20 iterations and 10 subsets was used. Reconstructed images were percentile-normalized, clipped to [0,1], and resized to the corresponding target dimensions when required.

The third pathway refined the OSEM image using a Residual U-Net. Instead of generating a complete image, the network predicted a correction:

$$\hat{x} = x_{OSEM} + R_\theta(x_{OSEM}) \tag{1}$$

where x_{OSEM} is the initial reconstruction and R_θ is the learned residual.

The model used convolutional encoder-decoder blocks with group normalization, skip connections, and bilinear decoder upsampling. The base channel width was reduced for the higher-resolution dataset to limit memory consumption. The network was trained with the following composite reconstruction objective:

$$\mathcal{L}_{\text{rec}} = \mathcal{L}_1(\hat{x}, x) + 0.25\mathcal{L}_{\text{MSE}}(\hat{x}, x) + 0.10\mathcal{L}_{\text{grad}}(\hat{x}, x) + 0.01\ \text{mean}(|R_\theta(x_{\text{OSEM}})|) \quad (2)$$

where x is the reference image. The $\mathcal{L}_1$ term provided overall intensity agreement, while the $\mathcal{L}_{\text{MSE}}$ term assigned a larger penalty to substantial reconstruction errors. The gradient term was defined as:

$$\mathcal{L}_{\text{grad}} = \mathcal{L}_1(\Delta_x\hat{x}, \Delta_x x) + \mathcal{L}_1(\Delta_y\hat{x}, \Delta_y x) \quad (3)$$

where Δ_x and Δ_y are first-order horizontal and vertical finite differences. This term encouraged preservation of local edges and uptake boundaries. The final regularization term constrained the magnitude of the predicted residual, encouraging the network to refine the OSEM reconstruction without introducing unnecessarily large corrections.

Both models were trained for 200 epochs using AdamW with learning rate 10^{-4} and weight decay 10^{-5}. Batch size was 16 for PPMI-PD Small and 1 for PPMI-PD Large because of the higher spatial resolution. Cosine annealing with a minimum learning rate of 10^{-5} was used for both datasets.

2.4 Classification Protocol

Four CNN backbones were evaluated: ResNet50, VGG16, Inception-v3, and Xception [14–17]. The same training protocol was applied across all backbones and input representations. A separate classifier was trained for each combination of CNN backbone, input representation, and diagnostic task.

All classifiers were trained for 500 epochs using cross-entropy loss, Adam optimization, batch size 16, learning rate 10^{-4}, and weight decay 10^{-5}. All classifiers were trained for 500 epochs using cross-entropy loss and the Adam optimizer. The batch size was 16, the initial learning rate was 10^{-4}, and weight decay was set to 10^{-5}. Cosine annealing with warm restarts was configured with $T_0 = 50$, $T_{\text{mult}} = 1$, and $\eta_{\min} = 10^{-6}$ [18]. The checkpoint with the highest validation accuracy was retained for evaluation.

Image-domain training included resizing, center cropping, random horizontal flipping, rotations within $\pm 10^{\circ}$, and brightness variation. Geometric augmentation was not applied to sinograms because rotation and flipping would change the relationship between projection angle and detector position. Sinograms were zero-padded to square dimensions before being passed to the CNNs. Direct sinogram classifiers were trained from random initialization.

Classification performance was evaluated using accuracy, AUROC, weighted precision, weighted recall, weighted F1-score, and expected calibration error [19]. For each task and input representation, Table 3 reports the highest test accuracy observed among the four evaluated CNN backbones. Detailed backbone-level metrics are provided separately.

Reconstruction quality was evaluated using PSNR and the global SSIM implementation used in the experimental code.

3 Results

3.1 Reconstruction Quality

Table 2 compares the three reconstruction pathways. OSEM followed by Residual U-Net refinement achieved the highest PSNR and global SSIM on both datasets.

Table 2. Reconstruction performance. Each entry reports PSNR in dB / global SSIM.

Dataset	Direct reconstruction	OSEM	OSEM + Residual U-Net
PPMI-PD Small	26.43 / 0.9606	25.17 / 0.9671	**39.62 / 0.9987**
PPMI-PD Large	24.63 / 0.9197	18.22 / 0.8302	**36.33 / 0.9934**

Direct reconstruction recovered the principal uptake structure but was less accurate than residual refinement. The difference was larger for the high-resolution dataset, where learning the complete non-square sinogram-to-image mapping was more difficult. Despite the acceptable quantitative performance of direct sinogram-to-image reconstruction, qualitative review by two physicians experienced in SPECT image interpretation identified alterations in several local uptake regions. These changes were not fully reflected by the global reconstruction metrics and could potentially affect clinical interpretation.

Residual refinement benefited from an OSEM initialization that already contained the approximate spatial distribution of tracer uptake. The neural network therefore needed to estimate only the remaining correction rather than the complete inverse transformation.

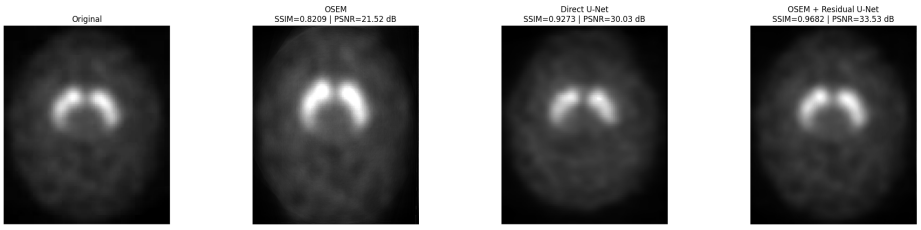

Fig. 2. Representative reconstruction comparison. Columns show the released reference image, direct AI reconstruction, OSEM reconstruction, and OSEM followed by Residual U-Net refinement. The selected example should illustrate visible smoothing or changes in local uptake that are not fully reflected by aggregate reconstruction metrics.

3.2 Diagnostic Performance

Table 3 summarizes the highest observed accuracy among the four CNN backbones for each pathway.

Table 3. Best observed classification accuracy among the four evaluated CNN backbones.

Task	Original image	Direct reconstruction	OSEM	OSEM + Residual U-Net	Direct sinogram
Small PD vs. non-PD	1.000	0.992	1.000	1.000	1.000
BigNPD	0.980	0.984	0.967	0.990	0.990
BigSWEDD	0.979	0.984	0.951	0.988	0.988
BigCONTROL	1.000	0.993	1.000	1.000	1.000

The Small and BigCONTROL tasks were close to saturation, and several representations reached perfect accuracy. Larger differences were observed for BigNPD and BigSWEDD.

OSEM alone produced the lowest best-observed accuracy in these two tasks. Residual refinement improved accuracy from 0.967 to 0.990 for BigNPD and from 0.951 to 0.988 for BigSWEDD. The increase in image similarity was therefore accompanied by an improvement in downstream classification.

Direct reconstruction retained sufficient diagnostic information despite its lower reconstruction metrics. Its best-observed accuracy reached 0.984 in both BigNPD and BigSWEDD.

Direct sinogram classification reached the same best-observed accuracy as OSEM + Residual U-Net in all four tasks. For BigSWEDD, the best direct-sinogram classifier achieved an AUROC of 0.999, compared with 0.984 for the best OSEM + Residual U-Net classifier. However, the test partition contained only 15 SWEDD images, and these differences should therefore be interpreted cautiously.

4 Discussion

The results demonstrate that reconstruction quality, automated diagnostic performance, and clinical image fidelity are related but not equivalent.

OSEM followed by Residual U-Net refinement produced the strongest image-domain representation. The residual formulation simplified the learning problem because the network received an approximate reconstruction and needed only to correct remaining local intensity and structural errors. This was particularly beneficial for the higher-resolution dataset.

Improved reconstruction quality also increased classification accuracy relative to OSEM alone in BigNPD and BigSWEDD. This confirms that reconstruction choices can affect downstream diagnosis and should not be evaluated only as an isolated image-restoration step.

At the same time, direct sinogram classification reached the same maximum accuracy as the refined reconstruction pathway. This suggests that the image-derived sinograms preserved discriminative information that could be used without first converting it into a conventional image. A CNN is not restricted to the representation that is easiest for a physician to interpret.

Importantly, acceptable global reconstruction metrics and successful downstream classification did not guarantee clinical fidelity of the directly reconstructed images. Qualitative review by two physicians experienced in SPECT image interpretation identified alterations in several local uptake regions that were not fully reflected by PSNR or SSIM. Because such distortions may affect the visual assessment of striatal uptake, the directly reconstructed images should not be considered suitable for standalone clinical interpretation in their current form. This observation indicates that reconstruction methods should be evaluated not only using global image-quality metrics and automated classification performance, but also through expert assessment of clinically relevant local structures.

These findings do not imply that clinical reconstruction is unnecessary. Reconstructed images remain essential for physician review, localization of abnormal uptake, quality control, and compatibility with established nuclear medicine workflows. Direct sinogram analysis should instead be considered a complementary machine-analysis pathway.

The strongest evidence comes from BigNPD and BigSWEDD because the Small and PD-versus-control tasks were almost saturated. Nevertheless, the SWEDD test subset was small, and one or two predictions could substantially affect the reported accuracy.

Several limitations remain. The sinograms were generated from already reconstructed images and therefore do not represent raw scanner measurements. The study used two-dimensional datasets prepared using different source preprocessing procedures. External validation was not performed. In addition, the main classification table reports the best result among four backbones rather than a repeated fixed-backbone comparison. Future work should evaluate native projection data, use independent validation and test partitions, report confidence intervals across repeated runs, and include external or multi-center cohorts.

5 Conclusion

This study evaluated DaT-SPECT reconstruction as part of a downstream diagnostic pipeline rather than as an isolated image-generation problem. OSEM followed by Residual U-Net refinement provided the strongest image-domain reconstruction and improved classification compared with unrefined OSEM. At the same time, classifiers trained directly on image-derived sinograms reached the same best-observed task-level accuracy as the hybrid reconstruction pathway.

The results do not support replacing reconstruction with a single reconstruction-free classifier. Instead, they show that the representations required for human interpretation and machine prediction may differ. Future SPECT systems may benefit from combining an interpretable reconstruction pathway with direct analysis of native projection-domain data, followed by validation on raw clinical acquisitions and clinically relevant diagnostic tasks.

References

1. Poewe, W., Seppi, K., Tanner, C.M., et al.: Parkinson disease. Nat. Rev. Dis. Primers **3**, 17013 (2017)
2. Akdemir, Ü.Ö., Bora, H.A.T., Atay, L.Ö.: Dopamine transporter SPECT imaging in Parkinson's disease and parkinsonian disorders. Turkish J. Med. Sci. **51**, 400–410 (2021)
3. Hudson, H.M., Larkin, R.S.: Accelerated image reconstruction using ordered subsets of projection data. IEEE TMI **13**, 601–609 (1994)
4. Shao, W., Rowe, S.P., Du, Y.: SPECTnet: a deep learning neural network for SPECT image reconstruction. Ann. Transl. Med. **9**, 819 (2021)
5. Kandarpa, V.S.S., Bousse, A., Benoit, D., et al.: DUG-RECON: a framework for direct image reconstruction using convolutional generative networks. IEEE TRPMS **5**, 44–53 (2021)
6. Yu, Z., Rahman, M.A., Laforest, R., et al.: Need for objective task-based evaluation of deep learning-based denoising methods: a study in the context of myocardial perfusion SPECT. Med. Phys. **50**, 4122–4137 (2023)
7. Rahman, M.A., Yu, Z., Laforest, R., et al.: DEMIST: a deep-learning-based detection-task-specific denoising approach for myocardial perfusion SPECT. IEEE TRPMS **8**, 439–450 (2024)
8. Kurmi, A., Biswas, S., Sen, S., et al.: An ensemble of CNN models for Parkinson's disease detection using DaTscan images. Diagnostics **12**, 1173 (2022)
9. Choi, H., Ha, S., Im, H.J., et al.: Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. NeuroImage: Clinical **16**, 586–594 (2017)
10. Wenzel, M., Milletari, F., Krüger, J., et al.: Automatic classification of dopamine transporter SPECT: deep convolutional neural networks can be trained to be robust with respect to variable image characteristics. Eur. J. Nucl. Med. Mol. Imaging **46**, 2800–2811 (2019)
11. Prashanth, R., Roy, S.D., Mandal, P.K., et al.: High-accuracy classification of Parkinson's disease through shape analysis and surface fitting in 123I-Ioflupane SPECT imaging. IEEE J. Biomed. Health Inform. **21**, 794–802 (2017)
12. Chegodaev, D., Handayani, L., Steven, R., et al.: Revisiting CNN-based Parkinson's disease classification from DaT-SPECT images: the role of training protocols. Electronics **15**, 1883 (2026)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
15. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2015)
16. Szegedy, C., Vanhoucke, V., Ioffe, S., et al.: Rethinking the Inception architecture for computer vision. In: CVPR. pp. 2818–2826 (2016)
17. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: CVPR. pp. 1251–1258 (2017)
18. Loshchilov, I., Hutter, F.: SGDR: stochastic gradient descent with warm restarts. In: ICLR (2017)
19. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: ICML. pp. 1321–1330 (2017)