Artificial Intelligence-Based Assistance in Clinical ^{123}I-FP-CIT SPECT Scan Interpretation

Evander van Wolfswinkel
Saint Antonius: Sint Antonius Ziekenhuis
Jette Wielaard
Sint Antonius Hospital: Sint Antonius Ziekenhuis
Jules Lavalaye
Saint Antonius: Sint Antonius Ziekenhuis
Jorrit Hoff
Saint Antonius: Sint Antonius Ziekenhuis
jan Booij
AMC-UvA Geneeskunde: Universiteit van Amsterdam Faculteit Geneeskunde
Tim de wit
AMC-UvA Geneeskunde: Universiteit van Amsterdam Faculteit Geneeskunde
Jan Habraken (j.habraken@antoniusziekenhuis.nl)
Saint Antonius: Sint Antonius Ziekenhuis

Research Article

Keywords: Brain, DaTSCAN, 123I-FP-CIT SPECT, Parkinson, AI, Machine Learning

Posted Date: July 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-721186/v1

Abstract

Purpose: Dopamine transporter (DAT) imaging with ^{123}I-FP-CIT SPECT is used to support the diagnosis of Parkinson's disease (PD) in clinically uncertain cases. Previous studies showed that automatic classification of ^{123}I-FP-CIT SPECT images (marketed as DaTSCAN) is feasible by using machine learning algorithms. However, these studies lacked sizable use of data from routine clinical practice. This study aims to contribute to the discussion whether artificial intelligence (AI) can be applied in clinical practice. Moreover, we investigated the need for hospital specific training data.

Methods: A convolutional neural network (CNN) named DaTNet-3 was designed and trained to classify DaTSCAN images as either normal or supportive of a dopaminergic deficit. Both a multi-site data set (n = 2412) from the Parkinson's Progression Marker Initiative (PPMI) and an in-house data set containing clinical images (n = 932) obtained in routine practice at the St Antonius hospital (STA) were used for training and testing. STA images were labeled based on interpretation by nuclear medicine physicians. To investigate whether indeterminate scans effects classification accuracy, a threshold was applied on the output probability.

Results: DaTNet-3 trained with STA data reached an accuracy of 89.0% in correctly identifying images of the clinical STA test set as either normal or with decreased striatal DAT binding (98.5% on the PPMI test set). When thresholded, accuracy increased to 95.7%. This increase was not observed when trained with PPMI data, indicating the incorrect images were confidently classified as the incorrect class.

Conclusion: Based on results of DaTNet-3 we conclude that automatic interpretation of DaTSCAN images with AI is feasible and robust. Further, we conclude DaTNet-3 performs slightly better when it is trained with hospital specific data. This difference increased when output probability was thresholded. Therefore we conclude that the usability of a data set increases if it contains indeterminate images.

Introduction

Parkinson's Disease (PD) is the fastest growing neurological condition in the world with prevalence rates increasing by about 74% from 1990 to 2016 [1]. PD is known for its distinct pathological changes, such as the degeneration of dopaminergic nigrostriatal neurons, projecting from the substantia nigra to the striatum of the brain. Early differentiation between patients with degeneration of dopaminergic neurons, and those without degeneration, is important for prognosis and treatment management. Dopamine transporter (DAT) imaging, using ^{123}I-FP-CIT single photon emission computed tomography (SPECT) (marketed as DaTSCAN), is currently the standard neuroimaging technique to support or exclude the diagnosis of dopaminergic deficit, consistent with PD and atypical parkinsonism, in clinically unclear cases. Varying inter-observer variability of human readers and the dependence on experienced nuclear medicine physicians make DaTSCAN [2, 3] interpretation an interesting task for artificial intelligence (AI) assisted classification. In literature several machine learning algorithms regarding this task are reported and show good performance [4–7]. However, there is a lack of studies evaluating the performance of

training and testing on sizable clinical data sets. Therefore, in this study we investigate and compare the need for data from clinical practice against a study data set aggregated from varying sources.

In this study we have thus designed a convolutional neural network (CNN) model for DaTSCAN interpretation. The model is trained with a publicly available data set and an in-house set of DaTSCAN images obtained in routine practice. In this study we investigated whether it is feasible to reliably classify DaTSCAN images using a CNN. Furthermore, we studied whether the model needs to be trained with camera and department specific data or whether a multi-site study data set can be used as a training set. Additionally, the effects of output probability on indeterminate images was studied.

Materials & Methods

DaTSCAN is a well-validated imaging tool used to investigate the loss of nigrostriatal dopaminergic neurons, by assessing DAT binding in the striatum. After injection of the radiotracer, SPECT imaging is performed, typically 3 to 4 h after injection [8], to create a 3D sliced representation of the striatal DAT binding.

Data

Parkinson's Progression Markers Initiative PPMI

A set of images was retrieved from the Parkinson's Progression Markers Initiative (PPMI) database [9]. This data set is used by many earlier DaTSCAN classifying studies [4–7] and is useful as a multi-site study data set from varying sources and for benchmarking. PPMI is a longitudinal study designed to assess the progression of PD using clinical features, biological markers and imaging data [9]. Acquisition protocols for the DaTSCAN imaging varied between originating centers. Yet, all centers used an 128 x 128 matrix, between 90 and 120 projections and an energy window centered on 159 +/- 10% KeV. Images were reconstructed using filtered back-projection or iterative reconstruction and are spatially normalized by registration to Montreal Neurological Institute (MNI) space by using PMOD (PMOD Technologies, Zurich, Switzerland) [9].

Images from the PPMI data set (dimension: 91 x 109 x 91, voxel size 2 x 2 x 2 mm^3) were further processed by extracting the binding region of the DAT-rich striatum which was assessed to be positioned in the same 20 slices in MNI space. Finally, images were downscaled to dimensions of 17 x 23 x 20.

This data set is referred to as the PPMI data set. The data set contains 351 normal control (NC) images originating from healthy controls and scans obtained in patients without evidence for dopaminergic deficit (SWEDD cohort) and 1422 dopaminergic deficit (DD) images originating from early PD patients.

Images obtained in routine practice

671 DaTSCAN images used in the differential diagnosis of clinically unclear patients were included retrospectively (December 2011 to February 2021) and these were acquired at the St. Antonius Hospital,

Nieuwegein. Data was acquired by the use of a double-head SPECT system (Siemens Symbia T2) with low-energy, high-resolution collimators. Scans were made 4 h after intravenous injection of ~ 185 MBq ^{123}I-FP-CIT, according to common guidelines [8, 10]. A total of 120 projections were acquired at 60 s per view for patients (128x128 matrix, zoom = 1). All SPECT images were reconstructed using 3D ordered-subsets-expectation-maximization (3D OSEM), using 4 iterations, 8 subsets and scatter correction. Reconstructions were filtered with an 8.4 mm Gaussian filter and CT-based attenuation correction was performed.

All images were spatially normalized by registration to MNI space by using SimpleITK [11]. The same processing steps were performed as used on the PPMI data set, resulting in final preprocessed images with the same dimensions and voxel size.

This data set is referred to as the St Antonius (STA) data set. Categorization of the clinical images into normal control (NC) (n = 377) and dopaminergic deficit (DD) (n = 294) was done based on the result of the original report by nuclear medicine physicians, using both visual and quantitative assessment. As this data set contains scans of clinically unclear patients and classifications are based on the interpretation reporting of single nuclear medicine physicians, it could contain incorrect classifications of indeterminate scans; such scans are referred to as indeterminates. The study protocol was examined by the Medical Research Ethics Committees United (institutional review board) of the St Antonius hospital, and they determined that, due to the nature of the research and since all patient data was fully anonymized, informed consent of the participants was waived. This study was conducted according to the Declaration of Helsinki.

Convolutional Neural Network Architecture

A CNN model, named DaTNet-3, was designed (summarized in *Fig. 1*), partly derived from the network by Mohammed et al [4], which in itself is a modified version of the AlexNet [12].

The DaTNet-3 architecture consists of three 3D convolutional layers, each followed by a max pooling layer (kernel size: 3 x 3 x 3, 128 feature maps). Feature maps are reduced by max pooling layers (taking the highest value of each feature map filter patch) for robustness [13]. In contrast to previous DaTSCAN classifying networks [4, 5], which normalize data in preprocessing, DaTNet-3 uses three batch normalization layers. This allows the distinguishable image feature to come more forward and lessen the effects of differing imaging sources [14]. Rectified linear units (ReLu) activations layers were added for each of the three layers, providing sensitive neuron activation and a lower computational cost, but avoiding easy saturation to a particular class [12]. Moreover, dropout was implemented in each of the three layers. Using dropout, noise is introduced to a part of the feature map inputs with a chance of 10% for additional regularization [15]. The final layer contains a global average pooling layer which down samples all feature maps to a single average value. This enforces a relation between previously generated feature maps and the output, allowing the interpretation of confidence for classifications [16]. The globally averaged value is fed into the output layer using a sigmoid function resulting in an output

between 0 and 1. Values under 0.5 are interpreted as NC while values above are considered images with DD. The closer the output is to either 0 or 1, the more likely the image belongs to the particular class.

DaTNet-3 training and setup

DaTNet-3 was trained and constructed using Tensorflow [17] through the Keras [18] interface. These open source libraries greatly simplify the coding, necessary to construct model architectures, training and benchmarking pipelines.

For training and testing DaTNet-3 models, both data sets were split into training and test sets. From both the PPMI and STA data sets a test set of 200 images was randomly retrieved with equal allocation of classes. Training images of the STA and PPMI data sets were randomly augmented to increase the amount of images, as machine learning models generalize better with larger data sets [19]. STA image count was doubled to 932 images (550 NC and 382 DD) using horizontal flips and randomization of intensity and brightness. Training images of the PPMI data set were similarly augmented, but only on the images obtained in the NC class to decrease the class imbalance. PPMI data set size was increased to 2412 images (1240 NC and 1172 DD).

Two DaTNet-3 models were generated and trained with different data input; DaTNet3_STA was trained using only training images of the STA data set and DaTNet-3_PPMI was trained using only PPMI training images. The results of DaTNet-3_PPMI on the PPMI test set were used to benchmark our model against similar DaTSCAN classifying studies. The results of DaTNet3_STA and DaTNet-3_PPMI on the STA test set were used to assess the usability in clinical practice and to investigative if models trained on multi-clinic study data can work well on data obtained in routine practice.

DaTNet-3 model performance parameters

To evaluate the models, accuracy, sensitivity and specificity were calculated. To visualize these metrics, confusion matrices were plotted. Confusion matrices show performance with a fixed threshold set at 0.5, allowing easy interpretation of accuracy, sensitivity and specificity. Because the STA data set contains indeterminate scans, which show little decrease in DAT binding and are therefore hard to interpret by nuclear medicine physicians, classifications can be inconclusive or only suggestive on the presence or absence of a DD. To investigate the effect these scans have on output, a probability threshold was implemented that filtered out indeterminate STA testing images with a threshold between 0.2–0.8.

Results

Two models (DaTNet-3_PPMI and DaTNet-3_STA) were trained with differing data set inputs and evaluated on their ability to correctly label DaTSCAN images based on the presence or absence of a DD. The models were tested on 2 testing data sets containing STA images for measuring the performance on data obtained in routine practice and PPMI images for benchmarking against other DaTSCAN classifiers.

Performance evaluation PPMI / clinical data

Accuracy, sensitivity and specificity of the DaTNet-3_PPMI model on the DaTNet-3_PPMI test set were 98.5%, 100% and 97%, respectively. In Table 1 these results are compared with previous DaTSCAN classifying studies.

Table 1
Comparative analysis with DaTSCAN classifying studies

Study	Method	Testing image count	Accuracy	Sensitivity	Specificity
This study (DaTNet-3_PPMI)	3D CNN (20 slices)	PPMI (n = 200)	98.5%	100%	97%
Ding et al [6] (2021)	Diffusion map + LLE (ensemble)	PPMI (n = 193)	98% (± 0.02)	96%	97%
Mohammed et al [4] (2021)	3D CNN (91 slices)	PPMI (n = 2723) (10-fold*)	99.34%	99.04%	99.63%
Choi et al [5] (2019)	3D CNN (91 slices)	PPMI (n = 75)	96.0%	94.2%	100%
Prashanth et al [7] (2014)	SVM (Striatal binding ratio)	PPMI (n = 674) (10-fold)	96.14%	96.55%	95.03%
*10-fold cross validation used for training and testing					

In Table 2, the performance of the two models on the clinical test data set is shown and confusion matrices are plotted in *Fig. 2* The presently presented DaTNet-3_STA model performs slightly better (accuracy 89%) than the DaTNet-3_PPMI model (accuracy 84%). Furthermore, the accuracy of the DaTNet-3_STA model is compared to the published interobserver agreement in Table 3.

Table 2
Performance results DaTNet-3

		Models	
Test set	Metric	DaTNet-3_PPMI	DaTNet-3_STA
STA	Accuracy	84.5%	89.0%
	Sensitivity	71.0%	87.0%
	Specificity	98.0%	91.0%

Table 3
Comparison interobserver agreement

Model/Study	Metric	
DaTNet-3_STA	**Accuracy**	89.0%
Tondeur et al [2]	**Interobserver agreement**	76% (range: 37–100%)
Booij et al [3]	**Kappa-coefficient**	0.74 – 0.93

To study whether the 22 (DaTNet-3_STA) or 31 (DaTNet-3_PPMI) STA test images (see Fig. 2) that were misclassified have a lower probability, thresholding was implemented by filtering out images with class probability between 0.2–0.8. The results can be seen in Table 4.

Table 4
Performance results DaTNet-3 thresholded output

		Models	
Test set	**Performance metric**	**DaTNet-3_PPMI**	**DaTNet-3_STA**
STA	**Accuracy** (Unthresholded)	84.5%	89.0%
	Accuracy (Thresholded)	85.8%	95.7%
	Sensitivity (Thresholded)	71.7%	93.8%
	Specificity (Thresholded)	100.0%	97.5%
	Images filtered/ Total images	20/200	36/200
	NC correct/ all certain NC	95/95	81/83
	DD correct/ all certain DD	61/85	76/81

As can be seen in Table 4, the accuracy increased after thresholding from 89.0–95.7% for the DaTNet-3_STA model. This means that images with an uncertain class probability (0.2 to 0.8) have a higher likelihood of being misclassified. This increase was not seen in Table 4 for the PPMI trained model, which shows very little increase in accuracy (84.5–85.8%).

Discussion

In this study we designed a CNN second reader for automatic classification of DaTSCAN images. The DaTNet-3_PPMI model was trained with multi-site study data of the PPMI data set, and the DaTNet-3_STA model with hospital specific data obtained in routine practice. The performance of the DaTNet-3_PPMI model was assessed on the PPMI-test set to benchmark performance to other DaTSCAN classifying studies. As is shown in Table 1, DaTNet-3_PPMI performed accurately in the classification of PPMI images, similarly to the results of the studies by Mohammed et al [4] and Ding et al [6]. The model

by Prashanth and co-workers [7] was outperformed by DaTNet-3. They used striatal binding ratio (SBR) values rather than the images themselves. These imaging features can be affected by changes in reconstruction and normalizations steps [20, 21], making it less robust than our DaTNet-3 CNN method. It is worth noticing that some studies used 10-fold cross validation [4, 7] and another study was hampered by a small testing sample size [5]. This could lead to an overestimation of the performance of these studies, which makes our conclusion that our DaTNet-3 model is at least on-par or even better than these studies and more reliable. Moreover, since our model is tested on a larger test set, and without 10-fold cross validation, we contribute to the evidence that AI can be potentially feasible for classifying DaTSCAN images.

To study whether an AI model can be successfully trained with multi-site study data, we assessed the performance DaTNet3_PPMI and DaTNet-3_STA on STA images obtained in routine practice. The results in Table 2 show that DaTNet-3 trained with PPMI images performs slightly less accurately (84.5%) compared to DaTNet-3 trained with clinical STA images (89.0%). However, it is noticed that the DaTNet-3_PPMI model has a very high specificity (98%), which implies that if the DaTNet-3_PPMI model predicts that a STA scan has normal DAT binding, and consequently does not support the clinical diagnosis of PD, this is correct in 98% of cases.

We further studied the effects on performance of both models on classifying scans as uncertain, by thresholding output probability. The clinically trained model, DaTNet-3_STA, showed a large increase in performance (from 89.0–95.7%). This indicates that incorrect classifications were substantially between the 0.2–0.8 uncertain probability threshold. These results signify the importance of training with data that contains more indeterminate scans, as this allowed the model to get better insights into such images.

The lack of accuracy increase using DaTNet-3_PPMI (from 84.5–85.8%) suggests that the model trained with PPMI images is incapable of classifying the indeterminate STA images with uncertain probability. Rather it classifies incorrect classification as confidently (meaning with a threshold under 0.2 or above 0.8) NC or with DD. The difference is probably explained by the fact that the PPMI data set merely contains manifest PD and normal images, in contrast to the STA data set that also contains also other parkinsonian syndromes and indeterminate images. Therefore, we conclude that the usability of a multi-site study data set as a training dataset can improve if indeterminate images are included as well, which is common practice in routine clinical studies.

The performance of DaTNet-3 can be compared to the published interobserver agreement of DaTSCAN studies by Tondeur et al [2] and Booij et al [3] (37–100% and 0.74–0.93 inter-observer agreement/k-coefficients for reader pairs, respectively). This comparison may indicate that DaTNet-3 shows high agreement in the interpretation made by the nuclear medicine physicians. Therefore DaTNet3 is potentially valuable as a second reader for reading DaTSCAN images in routine practice. Also, it is noticed that both the models have a very high specificity for classifying images as being normal,

therefore a potential application of such an AI tool could be to detect normal non-deficit images with high confidence.

In spite of the high accuracy, DaTNet-3 does not take patient age into account, which is useful for differentiating age-related from parkinsonian-related dopaminergic depletion [22]. Future studies could include patient's age as an extra parameter for model input to further improve its accuracy.

Furthermore, it has to be noticed that all our testing and training data originates from general purpose SPECT cameras. Images from dedicated brain scanners, like the InSPira HD SPECT system [23], might lead to diagnoses that are based on detailed local diminished uptake, as can be seen with such high-resolution systems. Preliminary results (data not shown) of our model using a test set of DaTSCAN images acquired on the InSPira HD SPECT system indeed showed that the model did not generalize well enough for data from a dedicated brain scanner. Therefore, in future studies we propose to set-up a multi-site study data set with data from both general and brain-dedicated systems to evaluate if it can perform accurately also on data obtained ion brain-dedicated systems.

Currently, one of the largest hurdles in AI is retrieving useable historic data from usually unstructured sources [24] and much time was spent on creating processed and correctly labeled training data. In this study, there is a lack of detailed and non-subjective DaTSCAN interpretation reporting, making it difficult to precisely classify images. In the future, the use of a predefined 5 stage degeneration scale [25] stored in a structured format could be considered to improve model input.

Conclusions

A CNN model, DaTNet-3, was designed and used for the automatic interpretation of DaTSCAN images. Based on the results of our model we conclude that automatic interpretation of DaTSCAN images with AI is feasible. DaTNet-3 may show great potential in increasing diagnostic confidence. By acting as an automated secondary reader, nuclear medicine physicians can get more confidence in their diagnosis without the need for complicated image feature selection tools. Further we conclude that our model performs slightly better when DaTSCAN images from clinical practice are used. This difference increased when the effects of indeterminate images in classification reporting is investigated through output probability thresholding. This allowed insights into probability of images that could be incorrect classifications. Therefore, we conclude that the usability of a PPMI study data set as an AI training-set increases if it contains indeterminate images.

Abbreviations

DAT dopamine transporter

PD Parkinson's disease

AI Artificial Intelligence

CNN Convolutional neural network

NC Normal control

DD Dopaminergic deficit

PPMI Parkinson's Progression Markers Initiative

STA St. Antonius data set

SPECT single-photon emission computed tomography

MNI Montreal Neurological Institute

SWEDD Scans without evidence of dopaminergic degeneration

3D OSEM 3D Ordered Subsets Expectation Maximization

METC Medical Ethical Testing Commission

ReLU Rectified Linear Units

SBR Striatal Binding Ratio

Declarations

Funding

Not applicable

Conflicts of interest/Competing interests

JB is consultant at GE healthcare (all payments paid to the institution)

All other authors declare that they have no conflict of interest.

Availability of data and material

All DaTSCAN images retrieved from the PPMI data set are available after application at https://www.ppmi-info.org/.

In-house STA DaTSCAN images are not available and confidential.

Code availability

All software packages used are open-source, derived/self-made software for processing/generating models not available.

Authors' contributions

EW participated in design of the study, wrote software implementation, designed the model, performed analysis and drafted the manuscript with help from JW, JB and JHA. JW and JHA participated in design of the study, supervised/conceptualized the project. JW performed the acquisition of data. JL and JHO helped interpret the data, provided domain knowledge and reviewed the manuscript. JB provided domain knowledge, critically reviewed/shaped the manuscript. TW participated in data acquisition and reviewed the manuscript.

All authors read and approved final manuscript.

References

1. Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2018;17:939–53. https://doi.org/10.1016/S1474-4422(18)30295-3
2. Tondeur MC, Hambye AS, Dethy S, Ham HR. Interobserver reproducibility of the interpretation of I-123 FP-CIT single-photon emission computed tomography. Nucl Med Commun. 2010;31:717–25. https://doi.org/10.1097/MNM.0b013e32833b7ea4
3. Booij J, Dubroff J, Pryma D, Yu J, Agarwal R, Lakhani P, et al. Diagnostic performance of the visual reading of123I-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
4. Mohammed F, He X, Lin Y. An easy-to-use deep-learning model for highly accurate diagnosis of Parkinson's disease using SPECT images. Comput Med Imaging Graph. 2021;87:101810. https://doi.org/10.1016/j.compmedimag.2020.101810
5. 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
6. Ding J-E, Chu C-H, Huang M-N Lo, Hsu C-C. Dopamine Transporter SPECT Image Classification for Neurodegenerative Parkinsonism via Diffusion Maps and Machine Learning Classifiers. 2021. https://arxiv.org/abs/2104.02066. Accessed May 2021.
7. Prashanth R, Dutta Roy S, Mandal PK, Ghosh S. Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging. Expert Syst Appl. 2014;41:3333–42. https://doi.org/10.1016/j.eswa.2013.11.031

8. Booij J, Hemelaar TG, Speelman JD, de Bruin K, Janssen AG, van Royen EA. One-day protocol for imaging of the nigrostriatal dopaminergic pathway in Parkinson's disease by [123I]FPCIT SPECT. J Nucl Med. 1999;40:753–61.
9. Marek K, Chowdhury S, Siderowf A, Lasch S, Coffey CS, Caspell-Garcia C, et al. The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort. Ann Clin Transl Neurol. 2018;5:1460–77. https://doi.org/10.1002/acn3.644
10. Morbelli S, Esposito G, Arbizu J, Barthel H, Boellaard R, Bohnen NI, et al. EANM practice guideline/SNMMI procedure standard for dopaminergic imaging in Parkinsonian syndromes 1.0. Eur J Nucl Med Mol Imaging. 2020;47:1885–912. https://doi.org/10.1007/s00259-020-04817-8
11. Beare R, Lowekamp B, Yaniv Z. Image segmentation, registration and characterization in R with simpleITK. J Stat Softw. American Statistical Association; 2018;86. https://doi.org/10.18637/jss.v086.i08
12. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Advances in neural information processing systems. 2012;25:1097-105. https://doi.org/10.1145/3065386
13. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. ICET. 2018. https://doi.org/ 10.1109/ICEngTechnol.2017.8308186
14. Bjorck J, Gomes C, Selman B, Weinberger KQ. Understanding Batch Normalization. Adv Neural Inf Process Syst. 2018. pp. 7694–705. http://arxiv.org/abs/1806.02375.
15. Park S, Kwak N. Analysis on the dropout effect in convolutional neural networks. Computer Vision – ACCV 2016. 2017. p. 189–204.
16. Lin M, Chen Q, Yan S. Network In Network. 2013. https://arxiv.org/abs/1312.4400.
17. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016. http://arxiv.org/abs/1603.04467.
18. Chollet F. Manning. Deep Learning with Python. 2017.
19. Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data. IEEE Intell Syst. 2009;24:8–12. https://doi.org/10.1109/MIS.2009.36
20. Brahim A, Ramírez J, Górriz JM, Khedher L, Salas-Gonzalez D. Erratum: Comparison between different intensity normalization methods in 123iioflupane imaging for the automatic detection of Parkinsonism. Plos one. 2015;10. https://doi.org/10.1371/journal.pone.013027
21. 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;4:1-16. https://doi.org/10.1186/s40658-017-0175-6
22. 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

23. Adriaanse SM, de Wit TC, Stam M, Verwer E, de Bruin KM, Booij J. Clinical evaluation of [123I]FP-CIT SPECT scans on the novel brain-dedicated InSPira HD SPECT system: a head-to-head comparison. EJNMMI Res. https://doi.org/2018;8. 10.1186/s13550-018-0436-y
24. Roh Y, Heo G, Euijong Whang S. A Survey on Data Collection for Machine Learning A Big Data-AI Integration Perspective. IEEE Transactions on Knowledge and Data Engineering. 2019. https://doi.org/10.1109/TKDE.2019.2946162
25. Kahraman D, Eggers C, Schicha H, Timmermann L, Schmidt M. Visual assessment of dopaminergic degeneration pattern in 123I-FP-CIT SPECT differentiates patients with atypical parkinsonian syndromes and idiopathic Parkinson's disease. J Neurol. 2012;259:251–60. https://doi.org/10.1055/s-0032-1301620

Figures

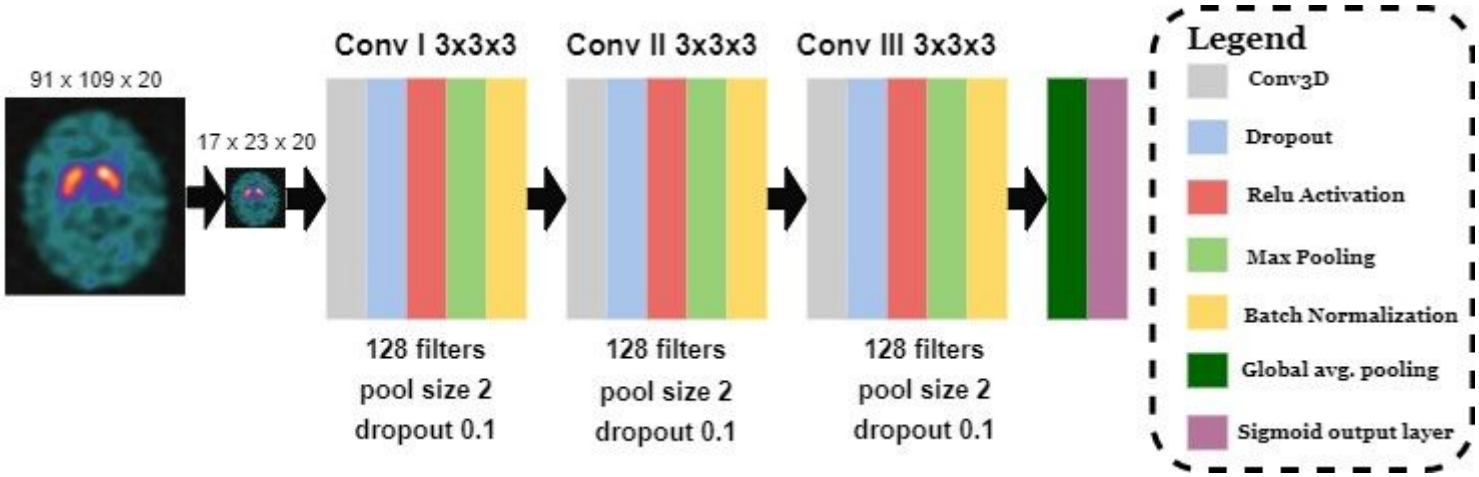

Figure 1

showing an abstract representation of DaTNet-3 architecture, allowing easy interpretation of the different layers and their order. Under each of the three section, it can be seen that layer parameters used stay the same for the entire model. Before training or inference begins, DaTSCAN images are reduced to dimensions of 17 x 23 x 20

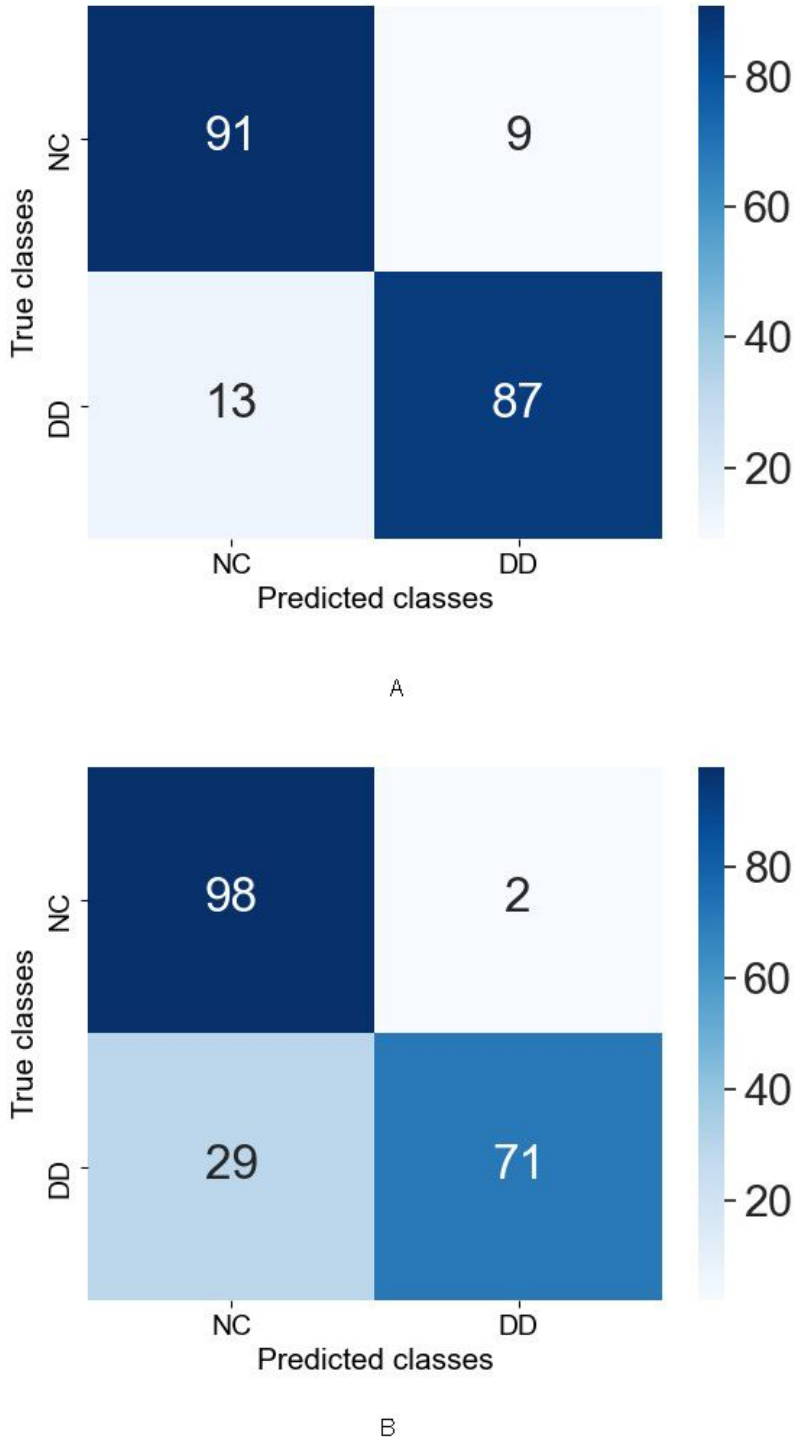

Figure 2

showing the confusion matrices for each model performance on STA test DaTSCANs. In these matrices, the correctly predicted normal control images are in the upper left, the falsely predicted normal control images in the upper right, the falsely predicted DD images in the lower left and the correctly predicted DD images in the lower right. 2a.) Confusion matrix showing the performance of DaTNet-3_STA on the STA testing data set. 2b.) Confusion matrix showing the performance of DaTNet-3_PPMI on the STA testing

data set. NC (Normal control image, image without visible dopaminergic deficit, but not necessarily healthy). DD (Dopaminergic Deficit, image with a visible dopaminergic deficit)