Clinical and Translational Imaging
https://doi.org/10.1007/s40336-026-00771-x

RESEARCH

From quantification to classification: comparative analysis of two software applications for machine learning–based prediction of early Parkinson's disease using ^{123}I-Ioflupane metrics

Luca Filippi[1,2] · Francesco Bianconi[3] · Andrea Marongiu[4] · Mario Luca Fravolini[3] · Matteo Minestrini[5] · Susanna Nuvoli[4] · Angela Spanu[4] · Orazio Schillaci[1] · Barbara Palumbo[5,6]

Received: 4 March 2026 / Accepted: 20 April 2026

Abstract

Purpose To compare semiquantitative striatal indices produced by two widely used quantification platforms — BasGanV2™ and NeuroTrans3D (Oasis©) — for ^{123}I-Ioflupane DAT SPECT, to evaluate their diagnostic performance for Parkinson's disease (PD) versus non-degenerative presentations, and to determine whether supervised machine-learning (ML) classifiers can clarify relative utility of the two methods.

Methods Retrospective analysis of 90 consecutive subjects (48 PD, 42 not-PD) undergoing routine ^{123}I-Ioflupane SPECT. Striatal binding ratios (caudate and putamen, left and right) were computed with both tools. Distributional differences were tested (Mann–Whitney U and Brown–Forsythe/Levene tests). Three supervised classifiers (Gaussian naïve Bayes, k-nearest neighbours, and SVM-RBF) were trained on four regional metrics from each tool using stratified shuffle-split (60:40 train: test) repeated 50 times; performance was summarised as mean (95% percentile CI) accuracy, sensitivity, specificity and AUC.

Results Both quantification methods discriminated PD from not-PD (all regions $p<0.0001$). NeuroTrans3D/Oasis© produced consistently higher mean binding values and significantly greater variance than BasGanV2™. ML models trained on either tool achieved excellent AUCs (0.917–0.960) and similar accuracy (≈86–91%), with overlapping confidence intervals and no single tool/classifier combination clearly outperforming the others.

Conclusion BasGanV2™ and NeuroTrans3D (Oasis©) are comparably effective for distinguishing PD from non-degenerative cases when combined with straightforward ML classifiers, despite systematic differences in absolute values and dispersion between tools. ML aids comparison by quantifying discriminative performance but, in this dataset, does not indicate a clear superiority of one quantification pipeline over the other.

Keywords ^{123}I-Ioflupane · Parkinson's disease · Dopamine transporters · BasGanV2 · Oasis© · Quantitative analysis · Neurology · Rehabilitation · Movement disorders · Personalized medicine

Luca Filippi M.D. and Francesco Bianconi contributed equally to this work.

✉ Luca Filippi
luca.filippi@uniroma2.it

[1] Department of Biomedicine and Prevention, University of Rome Tor Vergata, Via Montpellier, 1, Rome 00133, Italy

[2] Nuclear Medicine Research Unit, IRCCS San Raffaele Roma, Rome, Italy

[3] Department of Engineering, Università degli Studi di Perugia, Perugia, Italy

[4] Unit of Nuclear Medicine, Department of Medicine, Surgery and Pharmacy, University of Sassari, Sassari, Italy

[5] Nuclear Medicine Division, Azienda Ospedaliera di Perugia, S. Andrea delle Fratte, Perugia, Italy

[6] Section of Nuclear Medicine and Health Physics, Department of Medicine and Surgery, Università degli Studi di Perugia, Perugia, Italy

Published online: 06 May 2026

Introduction

The differential diagnosis of movement disorders—particularly between Parkinson's disease (PD) and non-degenerative tremor or parkinsonian conditions—remains primarily grounded in clinical assessment [1, 2]. Nevertheless, early-stage presentations frequently overlap, making diagnostic accuracy challenging and often requiring the integration of multimodal data. Age-related gait changes and tremor may represent a consequence of aging [3], but also early clinical symptoms preluding to a neurodegenerative condition, such as a parkinsonian syndrome. In this early phase, the diagnostic tools available are useful for neurologists and geriatricians to support the diagnostic process. In this context, dopamine transporter (DAT) single-photon emission computed tomography (SPECT) with ^{123}I-Ioflupane has become an established imaging biomarker for the in vivo evaluation of presynaptic dopaminergic integrity to contribute to a clear and early differential diagnosis among parkinsonian syndromes and normal aging conditions [4].

DAT SPECT enables visualisation of striatal dopamine transporter availability, thereby supporting the distinction between neurodegenerative parkinsonism, such as PD, and conditions not characterised by nigrostriatal degeneration. Although expert visual interpretation represents the clinical reference standard, qualitative assessment is inherently subjective and may be affected by inter-observer variability, particularly in patients with mild or equivocal findings [5].

To overcome these limitations, semiquantitative analysis methods have been developed to provide objective indices of tracer uptake within predefined striatal regions [6, 7]. These approaches typically generate striatal binding ratios (SBRs) for the caudate nuclei and putamina, improving diagnostic reproducibility and facilitating longitudinal and inter-centre comparisons. Among the available tools, Basal Ganglia V2 (BasGanV2™), a freely available software (http://www.aimn.it/struttura/gruppi/gs_neuro.php), and the dedicated neurological application incorporated into the commercially available Oasis© software (Segami Corporation) represent widely adopted platforms implementing automated or semi-automated region-of-interest (ROI) delineation and uptake quantification according to international procedural standards [8, 9].

Beyond quantification, the increasing availability of structured imaging metrics has opened new perspectives for computational analysis. Machine learning techniques—particularly supervised classifiers such as Support Vector Machines (SVMs)—have demonstrated strong performance in pattern recognition tasks using neuroimaging datasets [10]. By leveraging multidimensional semiquantitative descriptors, these algorithms may enhance diagnostic discrimination, especially in early disease stages where uptake reductions are subtle [9, 11]. In this scenario, the present study was designed around three primary endpoints: (1) to calculate and compare semiquantitative striatal uptake ratios obtained with BasGanV2™ and Oasis©; (2) to assess the diagnostic performance of the two methods in differentiating PD from non-degenerative conditions using clinical follow-up as reference standard; and (3) to evaluate the predictive value of software-derived metrics within a supervised machine learning model, including the relative contribution of caudate and putamina descriptors in early disease classification.

Materials and methods

Study design

This retrospective observational, single-centre study was conducted at the Nuclear Medicine Unit of the University Hospital of Sassari and included patients referred for ^{123}I-Ioflupane DAT SPECT imaging due to suspected mild parkinsonian symptoms, primarily bradykinesia and/or rigidity, with the clinical aim of confirming or excluding PD. The final diagnosis was established based on clinical assessment and neurological follow-up lasting at least 12 months, which served as the diagnostic reference standard [12].

Inclusion criteria were: (1) age ≥ 18 years; (2) clinical suspicion of PD; (3) availability of ^{123}I-Ioflupane SPECT imaging; and (4) a minimum clinical follow-up period of 12 months. Exclusion criteria comprised: (1) iodine allergy; (2) poor-quality or incomplete imaging data; and (3) uncertain or unstable clinical diagnosis after follow-up; and (4) diagnosis of atypical parkinsonian syndromes, including multiple system atrophy, progressive supranuclear palsy, and corticobasal degeneration. Furthermore, all medications known to potentially interfere with ^{123}I-Ioflupane uptake were discontinued at least 72 h prior to imaging, in accordance with international guidelines [13].

All patients underwent comprehensive clinical evaluation by neurology specialists. This assessment included routine laboratory investigations and brain MRI in order to rule out alternative diagnoses. Demographic and clinical variables — including sex, age, and disease duration — were systematically recorded. On the basis of follow-up outcomes, patients were classified into two groups: (1) PD — individuals with a clinically confirmed diagnosis of PD; and (2) not-PD — subjects in whom degenerative parkinsonism was excluded, as they were diagnosed as affected by age-related gait disturbances and essential tremor. For each participant, semiquantitative analysis of DAT SPECT data was carried out using two software packages, BasGanV2™

and Oasis©. For patients diagnosed with PD, further clinical data were collected, specifically the Hoehn and Yahr (H–Y) stage and scores on the Movement Disorder Society–Unified Parkinson's Disease Rating Scale (MDS-UPDRS). All clinical assessments were conducted during the "on" phase of the disease [14].

This retrospective study was performed for clinical purposes, in compliance with Institutional Review Board regulations and in accordance with the Declaration of Helsinki. As the analysis was based on data obtained during routine clinical care, formal approval from the institutional ethics committee was waived. Nevertheless, written informed consent for both the diagnostic procedures and the anonymised use of clinical data for scientific research was obtained from all patients prior to examination.

SPECT acquisition protocol

All patients underwent brain DAT SPECT imaging according to international procedural standards and European Association of Nuclear Medicine's practice guidelines [13]. Following thyroid blockade with oral potassium perchlorate (400 mg), ^{123}I-Ioflupane was administered intravenously at a target activity of 185 MBq, and image acquisition was performed approximately 3.5 h post-injection. SPECT studies were acquired using a dual-head gamma camera equipped with low-energy high-resolution collimators (Millenium MG, GE MEDICAL SYSTEMS), with the energy peak centred at 159 keV and a ± 10% window. Acquisition parameters included 180° rotation per detector, a 128 × 128 matrix, zoom factor 1.0, angular step of 3°, and 30 s per projection.

Patients were positioned supine, with the head immobilised using a dedicated support to minimise motion artefacts. Projection data were reconstructed using filtered back-projection with a Butterworth filter (cut-off frequency 0.5 cycles/cm; order 10). Reconstructed datasets were reoriented along the orbitomeatal line and reformatted into transaxial, coronal and sagittal planes with a slice thickness of approximately 2.2 mm. All scans also underwent expert visual assessment by experienced nuclear medicine physicians blinded to clinical data, for qualitative classification as normal or abnormal based on striatal uptake patterns.

Image processing

Image analysis was performed using two dedicated software platforms, BasGanV2™ and NeuroTrans3D for each enrolled subject. For both quantitative tools (BasGanV2™ and NeuroTrans 3D/Oasis©), image processing was performed by two nuclear medicine physicians (A.M. and S.N.), each with more than five years of experience in neuroimaging. The assessments were conducted jointly, and discrepancies were resolved by consensus.

BasGanV2™ (Supplemental data, figure S1) utilises a high-definition, three-dimensional striatal template based on the Talairach atlas, as previously reported [15]. The software enables fully automated 3-D segmentation of the caudate nucleus and putamen in both hemispheres. An optimisation procedure automatically refines the alignment of smoothed templates to achieve the best fit with the radioactive signal distribution and simultaneously positions an occipital ROI for background assessment. Correction for the partial volume effect (PVE) is incorporated during the calculation of binding values for the caudate nucleus, putamen, and background (occipital cortex). This correction follows a Talairach–Tournoux atlas-based three-compartment model of the basal ganglia, in which activity is allocated accordingly. Binding values for the putamen and caudate nucleus are adjusted by subtracting background binding using the formula:

$$[(caudate\ nucleus\ or\ putamen\ binding - background\ binding)/background\ binding]$$

This yields the SBR value for the caudate nucleus and putamen in each hemisphere.

The neurological tool NeuroTrans 3D (Supplemental data, figure S2), was implemented within the Mirage Processing Application (Segami Corporation, version 5.5), now integrated into the Oasis© platform. This software characterises striatal dopaminergic activity in terms of binding potentials (BP), incorporating attenuation correction and partial volume effect correction through a deformable three-dimensional model segmentation derived from the modified Talairach atlas. A BP threshold of 3.3 for both the caudate nucleus and putamen was established based on 20 healthy controls matched for sex and age (mean values: 4.9 ± 0.71 for the caudate and 4.6 ± 0.67 for the putamen) [8].

Although NeuroTrans3D/Oasis© provides quantitative measures expressed as BP and BasGanV2™ reports SBR, both metrics were considered equivalent indices of dopaminergic activity, as they reflect striatal DAT availability. For the sake of clarity and consistency, the sole term "SBR" is used throughout the manuscript as a general descriptor of dopaminergic activity derived from both software platforms.

Statistical analysis

Statistical differences in the SBR between PD vs. not-PD and between BasGanV2™ vs. NeuroTrans 3D (Oasis©) were assessed through non-parametric two-sided Mann-Whitney U test. Differences in the SBR variance between the two tools were evaluated by Levene test with the data

centred around the median (Brown-Forsythe test). All the results were considered significant for p-values < 0.05 after Bonferroni's correction for multiple tests.

All the analyses were carried out using Python 3.14.0 with functions from NumPy 2.3.5, pandas 2.3.3 and SciPy 1.16.3. Data visualisation was based on the same version of Python plus Matplotlib 3.10.8, Seaborn 0.3.12 and statannotations 0.7.2 [16].

Machine learning analysis

Supervised classification experiments were carried out to assess the ability of the SBR values from the four anatomical regions (CN_L, CN_R, PUT_L and PUT_R) to differentiate PD vs. not-PD subjects. Classifiers were trained using the features extracted either through BasGanV2™ or NeuroTrans3D (Oasis©). Three classification models were considered: Gaussian Naïve Bayes (GNB), k-Nearest Neighbours (k-NN) and Support Vector Classifier with radial-basis function kernel (SVM). For the k-NN and SVM grid-search hyperparameter tuning was also performed by 3-fold cross validation on the train set at each train/test split (as described below). We exhaustively searched the number of neighbours k ∈ {1,3,5} for the k-NN and the value of the penalization parameter C ∈ {0.01,0.1,1.0,10.0} for the SVM. For all the other parameters the default settings provided by scikit-learn were used.

Accuracy estimation was based on stratified shuffle split with a train ratio of 0.6—that is, at each split 60% of the subjects were used to train the classification models and the remaining 40% to estimate the accuracy. Stratification guaranteed that the PD vs. not-PD ratio in the train and test sets was the same as in the whole population. The process was repeated for $n = 50$ random splits into train and test sets (Supplementary material.splits.csv), and four performance metrics (accuracy, sensitivity, specificity and AUC) were recorded at each split. The pooled results were presented as mean and 95% confidence interval (CIs), the latter being estimated as the 2.5th and 97.5th percentile of the distribution of each metric over all the random splits.

At each train/test split preliminary z-score feature normalization was carried out on the train and test set separately, with the scaler always fitted on the train data only to prevent data leakage. The analysis was carried out using Python, NumPy and Pandas (see above for versions) plus scikit-learn 1.8.0.

Table 1 Demographic and clinical characteristics of patients and controls

Variable	PD ($n=48$)	Not-PD ($n=42$)	Total ($n=90$)
Age, years (mean ± SD)	66.1 ± 10.4	69.3 ± 9.2	67.6 ± 9.9
Age range, years	39–81	37–82	37–82
Male/Female (n)	35/13	23/19	58/32
Disease duration, months (mean ± SD)	11.2 ± 4.6	–	–
Hoehn and Yahr stage (mean ± SD)	1.6 ± 0.5	–	–
UPDRS part III score (mean ± SD)	18.3 ± 6.7	–	–
De novo, drug-naïve at imaging (n, %)	48 (100%)	–	–
Follow-up duration, months (mean ± SD)	18.4 ± 2.1	18.1 ± 2.4	18.3 ± 2.2

PD: Parkinson's disease; UPDRS: Unified Parkinson's Disease Rating Scale

Results

Demographic and clinical characteristics

A total of 90 consecutive patients presenting with mild parkinsonian features—predominantly subtle tremor, bradykinesia and rigidity—were included in this retrospective study and underwent ^{123}I-Ioflupane SPECT brain imaging to support diagnostic clarification. Following a minimum clinical follow-up of 12 months, 48 patients were diagnosed as PD and 42 as not-PD (subjects affected by normal aging brain + unsteady gait and subjects affected by Essential Tremor). Patients with PD were in an early stage of the disease and also de novo and drug-naïve at the time of DAT imaging. The overall cohort comprised 58 men and 32 women, with age of 67.6 ± 9.9 years (range 37–82 years). The PD group included 35 men and 13 women, with a mean age of 66.1 ± 10.4 years, while the not-PD group included 23 men and 19 women, with a mean age of 69.3 ± 9.2 years. In the PD group, mean disease duration at the time of SPECT examination was 11.2 ± 4.6 months, and the mean Hoehn and Yahr stage was 1.6 ± 0.5, consistent with mild, early disease. The mean UPDRS (motor examination) score was 18.3 ± 6.7. No significant age difference was observed between the two groups, and none of the control subjects developed clinical signs of degenerative parkinsonism during follow-up. Patients' characteristics are summarized in Table 1, and the anonymized data were provided as supplementary information (Supplementary material.dataset.csv).

PD vs. not-PD

At Mann-Whitney U test, the PD cases had significantly lower SBR than the not-PD in all the four anatomical regions, as one would reasonably expect (Table 2; Figs. 1—**2**). The results were consistent across the two quantification tools.

BasGanV2™ vs. NeuroTrans 3D (Oasis©)

NeuroTrans 3D (Oasis©) yielded significantly higher SBR than generated by BasGanV2™ in all the four anatomical regions, with mean values respectively in the range 3.66—5.45 and 2.49—3.85 (Table 3; Fig. 3). On Levene test, NeuroTrans 3D (Oasis©) also resulted in significantly larger data variance (dispersion) than BasGanV2™, with standard deviation respectively in the range 1.62—1.85 and 1.01—1.12. This result was again consistent across the four anatomical regions considered (Table 3; Fig. 3).

Machine learning–based prediction of Parkinson's disease

The ability to distinguish PD vs. not-PD cases was shown to be generally good by machine learning analysis based on GNB, kNN, and SVM classification models trained on SBR values from all four anatomical locations (CN_L, CN_R, PUT_L, and PUT_R). Based on Çorbacıoğlu and Askel's scale [17] the outcome was excellent in terms of AUC, with values in the range 0.917—0.960 (Table 4; Figs. 4—**5**). The other metrics also attested good performance in terms of accuracy, sensitivity and specificity with values ranging from 85.8% to 91.2%, 83.8% to 89.3%, and 88.0% to 97.6%, respectively (Table 4). The confidence intervals for all the performance metrics indicated substantial overlap across the classifiers (GNB, kNN and SVM) and quantification tools used (BasGanV2™, NeuroTrans3D/Oasis©) with no combination emerging as a clear winner (Table 4). Likewise, the ROC curves showed a comparable trend across the classification models and quantification tools (Figs. 4—**5**).

Table 2 Results of the univariate analysis: PD vs. not PD

Region	Tool	PD	not PD	*p*-value
CN_L	BasGanV2	3.32±0.88	4.45±0.79	<0.001*
CN_R	BasGanV2	3.17±0.87	4.49±0.68	<0.001*
PUT_L	BasGanV2	1.77±0.68	3.55±0.71	<0.001*
PUT_R	BasGanV2	1.71±0.69	3.37±0.62	<0.001*
CN_L	NeuroTrans3D (Oasis)	4.59±1.33	6.43±1.50	<0.001*
CN_R	NeuroTrans3D (Oasis)	4.32±1.35	6.32±1.19	<0.001*
PUT_L	NeuroTrans3D (Oasis)	2.57±1.20	5.28±1.33	<0.001*
PUT_R	NeuroTrans3D (Oasis)	2.30±1.04	5.20±1.18	<0.001*

Striatal binding ratios (PD, not-PD) are presented as mean±std, *p*-Value tests differences between the two groups. Asterisks indicate significant differences

Discussion

The present study provides a structured comparison between two semiquantitative pipelines for ^{123}I-Ioflupane DAT SPECT analysis, namely BasGanV2™ and NeuroTrans3D (Oasis©), and examines whether supervised machine-learning approaches contribute to clarifying their

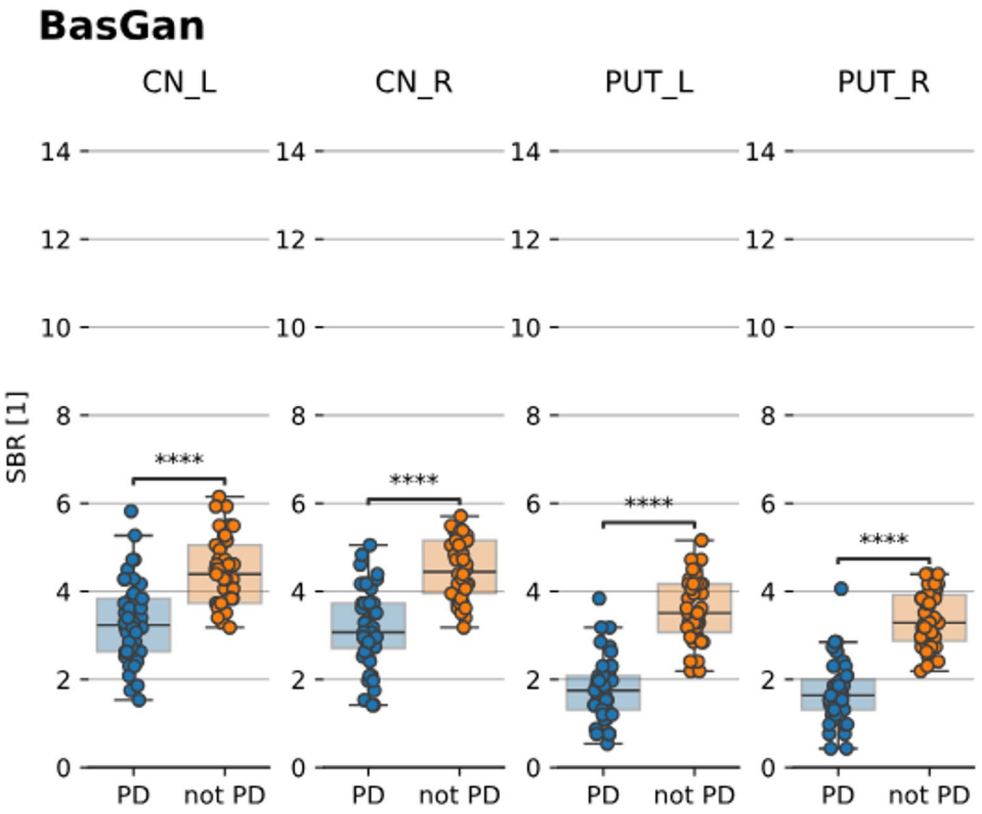

Fig. 1 Box plots and strip plots showing the SBR distribution of PD vs. not-PD in each anatomical region using BasGanV2™ as the quantification method. Each dot represents one patient; asterisks indicate the significance level of pairwise differences ('****': p-value < 0.0001)

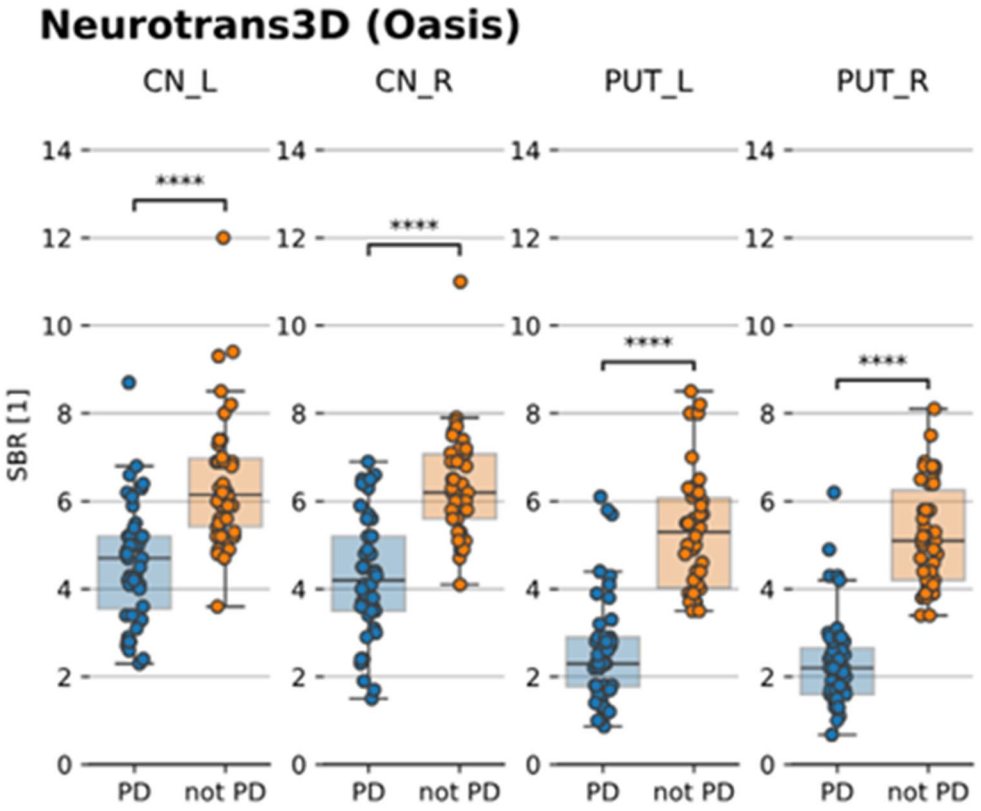

Fig. 2 Box plots and strip plots showing the SBR distribution of PD vs. not-PD in each anatomical region using NeuroTrans3D (Oasis©) as the quantification method. Each dot represents one patient; asterisks indicate the significance level of pairwise differences ('****': p-value < 0.0001)

Table 3 Results of the univariate analysis: BasGanV2™ vs. Neurotrans3D (Oasis)

Region	BasGanV2	NeuroTrans3D (Oasis)	*p*-value	*p*-value [var]
CN_L	3.85 ± 1.01	5.45 ± 1.68	< 0.001*	0.001*
CN_R	3.79 ± 1.03	5.25 ± 1.62	< 0.001*	< 0.001*
PUT_L	2.60 ± 1.12	3.84 ± 1.85	< 0.001*	< 0.001*
PUT_R	2.49 ± 1.06	3.66 ± 1.82	< 0.001*	< 0.001*

Striatal binding ratios are presented as mean ± std, *p*-Value tests differences between the two groups, *p*-value [var] differences in the variance between the two groups. Asterisks indicate significant differences

relative diagnostic performance in differentiating PD from non-degenerative conditions. Three principal findings emerge. First, both quantification systems demonstrated a robust ability to discriminate PD from not-PD across striatal subregions. This is of clinical relevance to contribute to a reliable diagnosis of movement disorders and age-related disturbances/essential tremor, thus strengthening the role of DAT SPECT quantification in this clinical challenge. Secondly, NeuroTrans3D (Oasis©) yielded consistently higher binding values and greater dispersion compared with BasGanV2™. Thirdly, when regional metrics derived from either platform were processed through established supervised classifiers, the diagnostic performance was generally good, with overlapping confidence intervals and no clear evidence of superiority of one quantification strategy over the other.

The observed differences in absolute values and variance are consistent with methodological distinctions between the two pipelines. NeuroTrans3D (Oasis©) employs deformable three-dimensional segmentation, attenuation correction and a binding potential framework calibrated against a reference population, whereas BasGanV2™ relies on a Talairach-template-based segmentation with atlas-derived partial volume correction and reports striatal binding ratios [18–20]. These divergent analytical frameworks are likely to influence scaling properties and measurement dispersion. The larger dynamic range observed with NeuroTrans3D (Oasis©) may reflect the combined impact of attenuation correction and model-based quantification, which can enhance contrast between specific and non-specific uptake. Conversely, the relatively compressed distribution produced by BasGanV2™ may indicate greater numerical stability across subjects.

Such distributional characteristics are not merely technical artefacts but may have practical implications. A wider dispersion may enhance separability near diagnostic thresholds, potentially favouring sensitivity in early or borderline cases, although it may also introduce greater variability in repeated measures. In contrast, a narrower distribution may support reproducibility and facilitate harmonisation across centres, particularly in longitudinal or multicentre settings

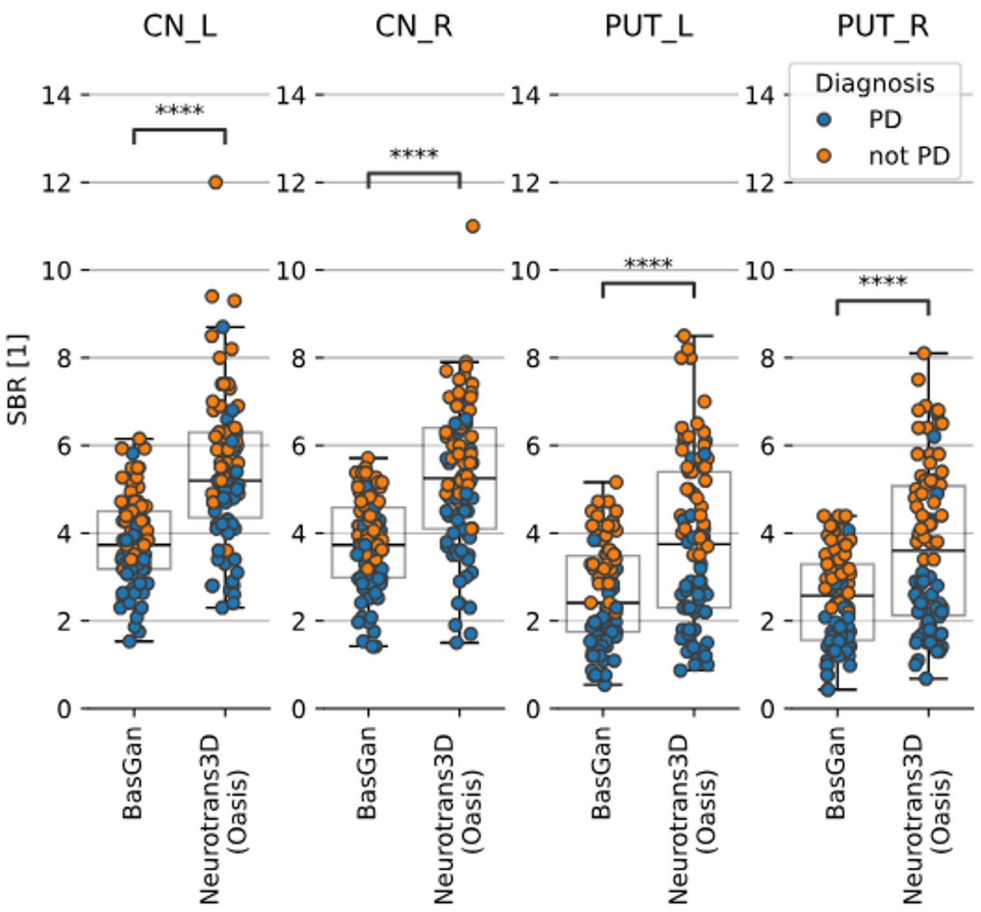

Fig. 3 Box plots and strip plots showing the SBR distribution obtained using BasGanV2™ vs. NeuroTrans3D (Oasis©) as the quantification method. Each dot represents one patient, colour denotes diagnosis (PD, not-PD) and asterisks indicate the significance level of pairwise differences ('****': p-value < 0.0001)

Table 4 Results of the machine learning analysis

Tool	Clf	Accuracy [%]	Sensitivity [%]	Specificity [%]	AUC
BasGanV2	GNB	85.8 [78.4–96.6]	83.8 [73.7–94.7]	88.0 [71.9–100.0]	0.944 [0.867–0.997]
BasGanV2	SVM	90.8 [81.2–99.4]	89.3 [73.7–100.0]	92.5 [76.5–100.0]	0.960 [0.910-1.000]
BasGanV2	kNN	88.9 [78.4–97.2]	87.2 [69.6–100.0]	90.9 [76.5–100.0]	0.917 [0.807–0.990]
NeuroTrans3D (Oasis)	GNB	87.4 [77.8–94.4]	87.3 [73.7–94.7]	87.6 [66.0-100.0]	0.943 [0.868–0.984]
NeuroTrans3D (Oasis)	SVM	91.2 [83.3–97.2]	85.5 [73.7–94.7]	97.6 [88.2–100.0]	0.947 [0.862–0.992]
NeuroTrans3D (Oasis)	kNN	90.7 [78.4–97.2]	85.4 [73.7–94.7]	96.7 [76.5–100.0]	0.917 [0.787–0.995]

Values are presented as mean [95% CI]. Key to abbreviations/acronyms: Clf = classifier, AUC = area under the curve;

where consistency is critical [18, 21]. From this perspective, the choice of quantification tool may reasonably be influenced by the clinical context, population characteristics and intended application, rather than by binary notions of superiority [22].

The integration of supervised machine-learning provides an additional layer of analysis. In this study, Gaussian naïve Bayes, k-nearest neighbours and support vector machines with radial basis kernels achieved excellent discrimination when trained on four regional features derived from either quantification platform. The similarity in performance across classifiers and tools indicates that the underlying biological signal of dopaminergic deficit in PD is sufficiently pronounced to be captured by diverse algorithmic approaches [22, 23]. Importantly, the application of machine-learning offers an objective framework for performance comparison, reducing reliance on isolated summary metrics and allowing repeated resampling to estimate variability. Nevertheless, within the constraints of the present dataset, machine-learning does not demonstrate a statistically meaningful advantage of one quantification method over the other.

The role of machine-learning in this context may therefore be conceptualised less as a mechanism for declaring a superior software solution and more as a tool for calibration,

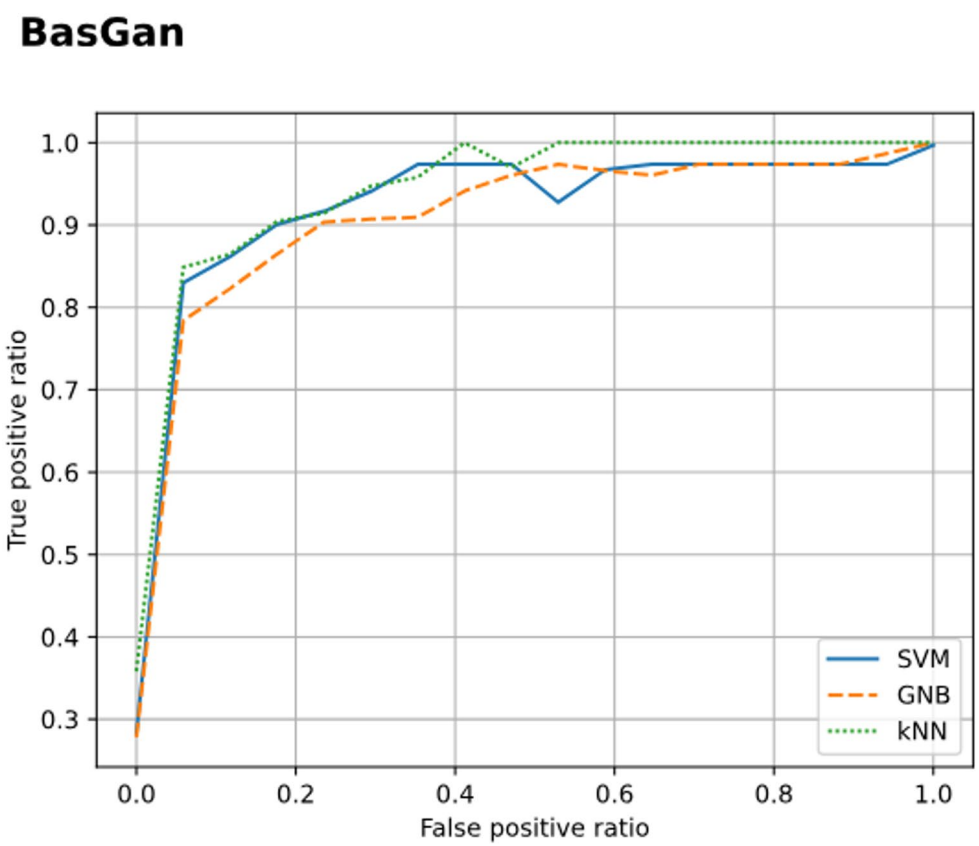

Fig. 4 Average receiver operating characteristic (ROC curves) showing the accuracy of the machine learning models trained on data generated using BasGanV2™ as the quantification model. The curves are averaged over all the train/test splits, colour and line type denote the classification model (GNB, kNN or SVM)

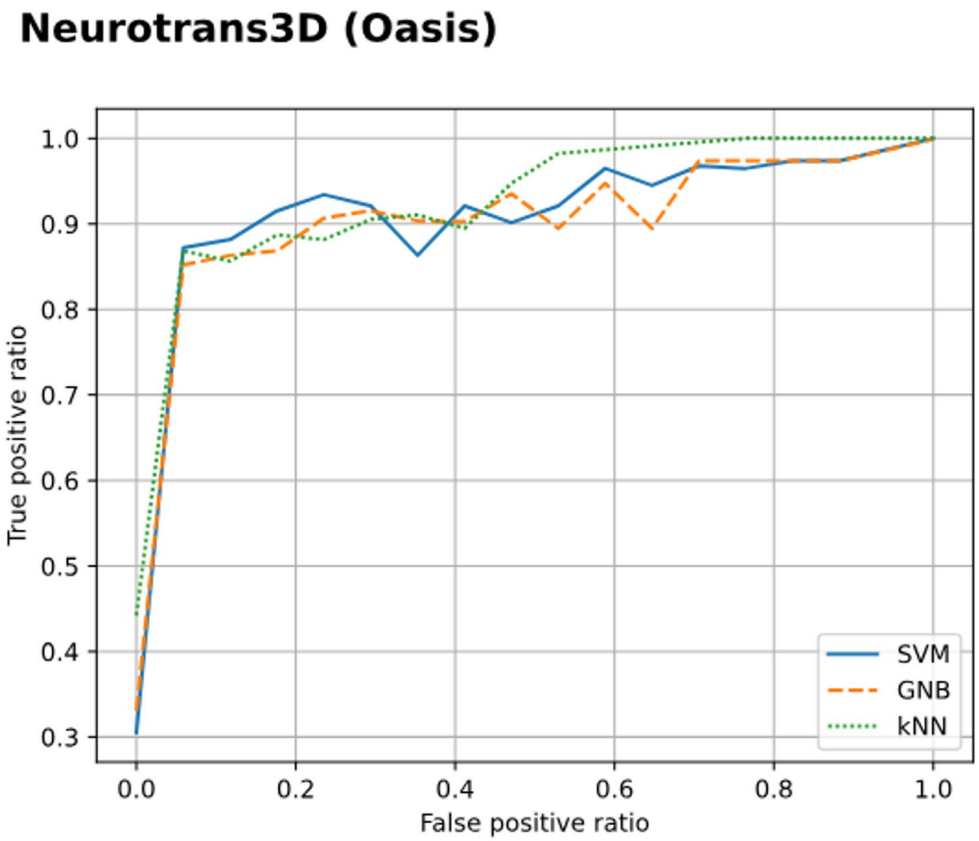

Fig. 5 Average receiver operating characteristic (ROC curves) showing the accuracy of the machine learning models trained on data generated using Neurotrans3D (Oasis©) as the quantification model. The curves are averaged over all the train/test splits, colour and line type denote the classification model (GNB, kNN or SVM)

optimisation and harmonisation [24]. By translating regional semiquantitative indices into probabilistic outputs, machine-learning models can support threshold selection tailored to local prevalence and clinical priorities [25, 26]. Furthermore, model-based approaches may facilitate cross-platform harmonisation, for example through regression-based mappings or ensemble strategies that integrate features from both pipelines. Such strategies could prove particularly relevant in multicentre studies or in healthcare systems where heterogeneous software infrastructures coexist.

Several limitations warrant consideration. This study has a single-centre design, which may limit the generalisability of the findings to other clinical settings. The retrospective nature and reliance on clinical follow-up as reference standard introduce the possibility of spectrum bias. Nevertheless, the study offers a novel and methodologically rigorous head-to-head comparison of two widely implemented semiquantitative pipelines, integrating conventional statistical analysis with supervised machine-learning to provide an original and clinically relevant perspective on their relative performance.

Conclusions

Both BasGanV2™ and NeuroTrans3D (Oasis©) provide reliable semiquantitative support for distinguishing PD from non-degenerative conditions on ^{123}I-Ioflupane DAT SPECT, thus contributing to the challenging differential diagnosis of movement disorders in the early ambiguous clinical phases. Although systematic differences in value distribution and variance were evident in our cohort, these did not result in meaningful differences in the classification performance. Our findings support a context-sensitive approach in which selection and calibration of quantification tools are aligned with the clinical population and intended use, complemented by algorithmic methods that enhance standardisation and interpretability.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40336-026-00771-x.

Acknowledgements None.

Author contributions Luca Filippi, Francesco Bianconi, Susanna Nuvoli and Barbara Palumbo: conceptualization, writing of the first draft and revision; Matteo Minestrini, Mario Luca Fravolini and Andrea Marongiu collection and interpretation of data; Orazio Schillaci and Angela Spanu revision of the first draft and supervision. All the authors read and approved the final version of the manuscript.

Funding Open access funding provided by Università degli Studi di Roma Tor Vergata within the CRUI-CARE Agreement. This work was partially supported by funding of the Italian Ministry of Health [Ricerca corrente].

Data availability Original data are available as supplementary material.

Declarations

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

Ethic approval This retrospective study was conducted for clinical purposes in accordance with institutional ethics committee regulations and the principles of the Declaration of Helsinki. As the analysis relied exclusively on routinely collected clinical data, formal approval by the institutional ethics committee was waived. All patients provided written informed consent prior to examination, including consent for diagnostic procedures and for the anonymised use of their clinical data for research purposes.

Consent for publication Informed consent was obtained by each participating subject. All data were anonymised to protect participant confidentiality.

References

1. Ali K, Morris HR (2015) Parkinson's disease: chameleons and mimics. Pract Neurol 15:14–25. https://doi.org/10.1136/practneurol-2014-000849
2. Jankovic J (2008) Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 79:368–376. https://doi.org/10.1136/jnnp.2007.131045
3. Nonnekes J, Post E, Imbalzano G, Bloem BR (2025) Gait changes with aging: an early warning sign for underlying pathology. J Neurol 272:257. https://doi.org/10.1007/s00415-025-12995-4
4. Brücke T, Brücke C (2022) Dopamine transporter (DAT) imaging in Parkinson's disease and related disorders. J Neural Transm (Vienna) 129:581–594. https://doi.org/10.1007/s00702-021-02452-7
5. Iwabuchi Y, Nakahara T, Kameyama M et al (2018) Quantitative evaluation of the tracer distribution in dopamine transporter SPECT for objective interpretation. Ann Nucl Med 32:363–371. https://doi.org/10.1007/s12149-018-1256-x
6. Neill M, Fisher JM, Brand C et al (2021) Practical Application of DaTQUANT with Optimal Threshold for Diagnostic Accuracy of Dopamine Transporter SPECT. Tomography 7:980–989. https://doi.org/10.3390/tomography7040081

7. Bailey DL, Willowson KP (2013) An evidence-based review of quantitative SPECT imaging and potential clinical applications. J Nucl Med 54:83–89. https://doi.org/10.2967/jnumed.112.111476
8. Nuvoli S, Spanu A, Piras MR et al (2017) 123I-ioflupane brain SPECT and 123I-MIBG cardiac planar scintigraphy combined use in uncertain parkinsonian disorders. Medicine 96:e6967. https://doi.org/10.1097/MD.0000000000006967
9. Nobili F, Naseri M, De Carli F et al (2013) 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 40:565–573. https://doi.org/10.1007/s00259-012-2304-8
10. Palumbo B, Bianconi F, Nuvoli S et al (2021) Artificial intelligence techniques support nuclear medicine modalities to improve the diagnosis of Parkinson's disease and Parkinsonian syndromes. Clin Transl Imaging 9:19–35. https://doi.org/10.1007/s40336-020-00404-x
11. Filippi L, Bianconi F, Frantellizzi V et al (2026) Machine learning for automated differentiation of parkinson's disease and its mimics using ^{123}I-mIBG scintigraphy: insights from a multicentre real-world cohort (ITA-mIBG study). Eur J Nucl Med Mol Imaging. https://doi.org/10.1007/s00259-025-07729-7
12. Postuma RB, Berg D, Stern M et al (2015) MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 30:1591–1601. https://doi.org/10.1002/mds.26424
13. Morbelli S, Esposito G, Arbizu J et al (2020) EANM practice guideline/SNMMI procedure standard for dopaminergic imaging in Parkinsonian syndromes 1.0. Eur J Nucl Med Mol Imaging 47:1885–1912. https://doi.org/10.1007/s00259-020-04817-8
14. Postuma RB, Poewe W, Litvan I et al (2018) Validation of the MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 33:1601–1608. https://doi.org/10.1002/mds.27362
15. Palumbo B, Filippi L, Marongiu A et al (2025) Diagnostic Accuracy of DaTQUANT® Versus BasGanV2™ for 123I-Ioflupane Brain SPECT: A Machine Learning-Based Differentiation of Parkinson's Disease and Essential Tremor. Biomedicines 13:2367. https://doi.org/10.3390/biomedicines13102367
16. Bianconi F (2024) Data and Process Visualisation for Graphic Communication: A Hands-on Approach with Python. Springer Nature Switzerland, Cham
17. Çorbacıoğlu ŞK, Aksel G (2023) Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. Turkish J Emerg Med 23:195–198. https://doi.org/10.4103/tjem.tjem_182_23
18. Kahraman D, Eggers C, Holstein A et al (2012) 123I-FP-CIT SPECT imaging of the dopaminergic state. Visual assessment of dopaminergic degeneration patterns reflects quantitative 2D operator-dependent and 3D operator-independent techniques. Nuklearmedizin 51:244–251. https://doi.org/10.3413/Nukmed-0449-11-12
19. Skanjeti A, Castellano G, Elia BO et al (2015) Multicenter Semiquantitative Evaluation of (123)I-FP-CIT Brain SPECT. J Neuroimaging 25:1023–1029. https://doi.org/10.1111/jon.12242
20. Skanjeti A, Angusti T, Iudicello M et al (2014) Assessing the accuracy and reproducibility of computer-assisted analysis of (123) I-FP-CIT SPECT using BasGan (V2). J Neuroimaging 24:257–265. https://doi.org/10.1111/jon.12008
21. Jovicich J, Barkhof F, Babiloni C et al (2019) Harmonization of neuroimaging biomarkers for neurodegenerative diseases: A survey in the imaging community of perceived barriers and suggested actions. Alzheimers Dement (Amst) 11:69–73. https://doi.org/10.1016/j.dadm.2018.11.005
22. Llera A, Huertas I, Mir P, Beckmann CF (2019) Quantitative Intensity Harmonization of Dopamine Transporter SPECT Images Using Gamma Mixture Models. Mol Imaging Biol 21:339–347. https://doi.org/10.1007/s11307-018-1217-8
23. Boyle AJ, Gaudet VC, Black SE et al (2021) Artificial intelligence for molecular neuroimaging. Ann Transl Med 9:822. https://doi.org/10.21037/atm-20-6220
24. Hu F, Chen AA, Horng H et al (2023) Image harmonization: A review of statistical and deep learning methods for removing batch effects and evaluation metrics for effective harmonization. NeuroImage 274:120125. https://doi.org/10.1016/j.neuroimage.2023.120125
25. Nenning K-H, Langs G (2022) Machine learning in neuroimaging: from research to clinical practice. Radiol (Heidelb) 62:1–10. https://doi.org/10.1007/s00117-022-01051-1
26. Singh NM, Harrod JB, Subramanian S et al (2022) How Machine Learning is Powering Neuroimaging to Improve Brain Health. Neuroinformatics 20:943–964. https://doi.org/10.1007/s12021-022-09572-9