Annals of Nuclear Medicine (2021) 35:378–385
https://doi.org/10.1007/s12149-021-01576-w

ORIGINAL ARTICLE

Clinical value of machine learning-based interpretation of I-123 FP-CIT scans to detect Parkinson's disease: a two-center study

M. Dotinga[1,5] · J. D. van Dijk[1] · B. N. Vendel[1,3] · C. H. Slump[5] · A. T. Portman[4] · J. A. van Dalen[2]

Received: 28 August 2020 / Accepted: 28 December 2020 / Published online: 20 January 2021

Abstract

Purpose Our aim was to develop and validate a machine learning (ML)-based approach for interpretation of I-123 FP-CIT SPECT scans to discriminate Parkinson's disease (PD) from non-PD and to determine its generalizability and clinical value in two centers.

Methods We retrospectively included 210 consecutive patients who underwent I-123 FP-CIT SPECT imaging and had a clinically confirmed diagnosis. Linear support vector machine (SVM) was used to build a classification model to discriminate PD from non-PD based on I-123-FP-CIT striatal uptake ratios, age and gender of 90 patients. The model was validated on unseen data from the same center where the model was developed ($n=40$) and consecutively on data from a different center ($n=80$). Prediction performance was assessed and compared to the scan interpretation by expert physicians.

Results Testing the derived SVM model on the unseen dataset ($n=40$) from the same center resulted in an accuracy of 95.0%, sensitivity of 96.0% and specificity of 93.3%. This was identical to the classification accuracy of nuclear medicine physicians. The model was generalizable towards the other center as prediction performance did not differ thereby obtaining an accuracy of 82.5%, sensitivity of 88.5% and specificity of 71.4% ($p=$NS). This was comparable to that of nuclear medicine physicians ($p=$NS).

Conclusion ML-based interpretation of I-123-FP-CIT scans results in accurate discrimination of PD from non-PD similar to visual assessment in both centers. The derived SVM model is therefore generalizable towards centers using comparable acquisition and image processing methods and implementation as diagnostic aid in clinical practice is encouraged.

Keywords I-123 FP-CIT · Machine learning · SPECT · SVM-model · Artificial intelligence

Introduction

Single positron emission computer tomography (SPECT) with I-123 N-ω-fluoropropyl 2β-carbomethoxy-3β-(4-iodophenyl)nortropane (FP-CIT) allows for visualization of striatal dopamine deficiency due to the loss of dopaminergic neurons which is characteristic of Parkinson's disease (PD) [1]. This imaging technique aids in the diagnostic process as it enhances diagnostic confidence, especially in patients with clinically uncertain parkinsonian syndromes [2, 3]. Re-evaluation of diagnosis is evident in up to 35% of these patients and changes in management and treatment are induced in approximately 70% [2, 4]. Current guidelines recommend the combination of visual assessment and semi-quantitative analysis for adequate interpretation of I-123 FP-CIT scans [5, 6]. Semi-quantitative analysis comprises the assessment of radiopharmaceutical-specific uptake in regions of interest (striatum, caudate nucleus and putamen) and non-specific uptake in reference areas as the occipital lobe. The addition of semi-quantification to visual assessment results in an increased reader confidence and superior diagnostic accuracy when compared to standalone visual assessment [7]. However, no universal cutoff values

✉ J. D. van Dijk
jorisvdijk@gmail.com

1 Department of Nuclear Medicine, Isala Hospital, PO Box 10400, 8000 GK Zwolle, The Netherlands

2 Department of Medical Physics, Isala Hospital, Zwolle, The Netherlands

3 Department of Nuclear Medicine, Treant Zorggroep, Emmen, The Netherlands

4 Department of Neurology, Treant Zorggroep, Emmen, The Netherlands

5 MIRA: Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands

are available to determine whether semi-quantitative results are normal or abnormal as acquisition, reconstruction and quantification methods are known to influence these striatal uptake ratios [8, 9].

A trend is currently seen towards machine learning (ML)-based approaches for automated classification of I-123 FP-CIT scans to improve interpretation [10, 11]. These approaches have the potential to be exploited as diagnostic aid to nuclear medicine physicians, thereby improving interpretation of these scans and overcoming the limitations of semi-quantitative analysis [10, 11]. Yet usage of developed models in clinical practice is limited as generalizability towards previously unseen data in different centers is often not determined. Assessment of the model's performance in independent validation datasets is necessary to ensure a model is valid for clinical use which is more and more recognized and addressed in the field. Widespread adoption of ML models requires a robust external validation and assessment of its clinical utility [12]. The aim of this study is therefore to develop a machine-learning based approach for interpretation of I-123 FP-CIT scans to detect PD and to determine its generalizability and clinical value in two centers using independent validation datasets from two centers.

Materials and methods

Study population

We retrospectively included a consecutive cohort of 210 patients with clinically confirmed diagnosis that underwent I-123 FP-CIT SPECT imaging between 2014 and 2018 in two medical centers in the Netherlands: Isala hospital, Zwolle (further referred by hospital 1) and Treant Zorggroep, Scheper hospital, Emmen (hospital 2). Patients' diagnoses were assessed by the attending neurologist according to standard diagnostic criteria. Clinical and demographic data including diagnosis, gender and age at the time of SPECT acquisition were retrieved from patients' medical records and patients were labelled as either having PD or a diagnosis other than PD (non-PD). The hospital 1 data was split into a group of 90 patients for model development (further referred by group A) and a group of 40 patients for validation (further referred by group B) using stratified random sampling. Data of the Scheper hospital ($n = 80$, further referred by group C) were used to assess the generalizability of the model. Both group B and C were used in the clinical evaluation. This study was retrospective and approval by the medical ethics committee was therefore not required according to the Dutch law. However, all procedures performed were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Image acquisition and reconstruction

Patients were instructed to discontinue medication interfering with I-123 FP-CIT binding to dopamine transporters prior to scanning. SPECT studies were carried out according to standard clinical procedure using two dual-headed gamma cameras (Infinia Hawkeye, GE Healthcare) equipped with a low energy, high resolution collimator. Three to six hours before SPECT acquisition, patients were intravenously administered 185 MBq of I-123 FP-CIT. A total of 64 projections over a circular 360° orbit (rotational radius of approximately 13 cm) were acquired on a 128×128 matrix (1.23–1.28 acquisition zoom, 3.45–3.59 mm pixel size) with an overall scanning time of 32 min (30 s per projection). A 10% energy window centered on the photopeak of I-123 at 159 keV was used.

Image reconstruction was performed by filtered back projection using a Butterworth pre-filter (cut-off 0.65 cycles/cm, order 10) and uniform Chang attenuation correction (coefficient 0.11 cm^{-1}). Attenuation correction was based on a variable ellipsoid map that followed the contour of the head, manually defined using thresholding. Images were reformatted into slices in axial, coronal and transversal planes (3.45–3.59 mm slice thickness) and axial slices were reoriented along the acanthomeatal line. The acquisition and reconstruction procedures and settings mentioned above were used in both centers.

Semi-quantitative analysis

Semi-quantitative analysis was performed using a functional imaging workstation (Xeleris version 4.0; GE Healthcare) to assess specific I-123 FP-CIT binding in the striatum and striatal sub-regions including both right and left caudate nucleus and putamen. Non-specific binding was assessed using the occipital cortex as reference region [6]. Five pre-defined fixed regions of interest (ROIs) were manually positioned over caudate nucleus, putamen and occipital cortex on three consecutive slices that were selected best representative of the activity and shape of the striatum. An example of a non-PD and PD patient with correct positioning of all predefined regions is shown in Fig. 1. The mean photon counts in each region was calculated over the three slices after which specific binding ratios (SBRs) were obtained according to the formula:

$$SBR = \frac{C_{str} - C_{occ}}{C_{occ}} \quad (1)$$

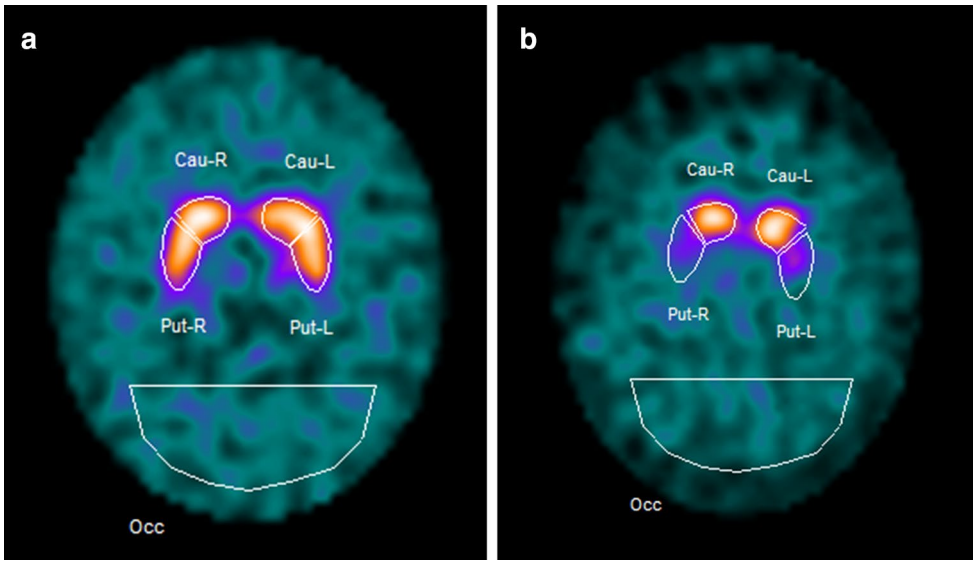

Fig. 1 Representative slice of a I-123 FP-CIT scan of a non-PD (**a**) and PD (**b**) patient with predefined regions for semi-quantitative analysis, including regions for the left and right caudate nuclei (Cau-L and Cau-R) and putamen (Put-R and Put-L) and the occipital cortex (Occ)

where C_{str} are the mean counts in the striatum or striatal sub-region and C_{occ} the mean counts in the occipital cortex. SBRs for both left and right striatum, caudate nucleus and putamen were assessed, resulting in six different ratios: Str_R, Str_L, $Caud_R$, $Caud_L$, Put_R and Put_L. Furthermore, a putamen/caudate index was calculated by dividing the mean photon counts within the putamen by the mean photon counts within the caudate nucleus, thereby obtaining ratios $Put_R/Caud_R$ and $Put_L/Caud_L$.

SVM classifier development and validation analysis

Support vector machine (SVM) is a widely used ML algorithm to build predictive models for classification problems. SVM attempts to find an optimal separation between classes by searching for a hyperplane or decision boundary that is able to separate the given data with minimal errors and that maximizes the margin between the classes. In this study, SVM was used as a classification method to discriminate PD from non-PD based on input features that included patient's age, gender and all eight I-123 FP-CIT striatal uptake ratios; Str_R, Str_L, $Caud_R$, $Caud_L$, Put_R, Put_L $Put_R/Caud_R$ and $Put_L/Caud_L$. Prior to training and validation procedures, ratios and age were normalized such that the mean value was 0 and standard deviation was 1. All procedures were performed using Matlab software (MATLAB and Statistics and Machine learning Toolbox Release 2018b, the Mathworks Inc.).

Group A was used to build a linear SVM and to perform hyperparameter optimization. A grid search was conducted, thereby evaluating regularization parameter values of 2^{-6}, 2^{-5}, 2^{-4}, …, 2^{8}. For each value, a stratified, 10-times repeated tenfold cross-validation was performed after which the mean F1-score was determined. The F1-score is defined as:

$$F_1 = 2 \cdot \frac{positive\,predictive\,value \cdot sensitivity}{positive\,predictive\,value + sensitivity} \quad (2)$$

The value providing the highest mean F1-score was used to derive the final model. This model was then validated in group B and C, assessing both predicted class and the probability that test data belongs to PD. For the latter, an inbuilt function for converting SVM scores to probabilities based on logistic regression was used. Given the probabilities, the validation datasets (group B and C) were divided into four categories (< 20%, 20–49%, 50–80%, > 80% probability of PD). Furthermore, prediction performance was determined by assessing F1-score, accuracy, sensitivity and specificity.

Clinical value

Whether and to what extent I-123 FP-CIT images and ratios were typical or characteristic for PD was assessed by two nuclear medicine physicians according to a 4-point scale. This scale consisted of the following categories: unlikely, not probable, probable and certain PD and comprised an expected probability of PD of < 20%, 20–49%, 50–80% and > 80%, respectively. Images were scored visually, taking into account the magnitude and homogeneity of I-123 FP-CIT distribution, striatal shape and symmetry, definition of striatal borders and the amount of background activity. Images were first scored without ratios after which ratios were presented to the nuclear medicine physician and a second score was obtained. In case of disagreement, overread from a third nuclear medicine physician was performed. Images were presented in random order and all readers were

blinded to patient characteristics and clinical information except age and gender. After scoring images of group B and C, accuracy, sensitivity and specificity to discriminate PD from non-PD were assessed for the nuclear medicine physicians and compared to the performance of the derived SVM model. Patients assigned to categories 1 or 2 (≤ 49% chance on PD) were categorized as non-PD and patients assigned to categories 3 and 4 (≥ 50% chance on PD) as PD.

Statistical analysis

Statistical analysis was performed using R Studio software [13]. To assess differences between group B and C in age, ratios and gender, the Mann–Whitney *U* test or χ^2-test were performed. Accuracies, sensitivities and specificities of the classifier for group B and C were compared using Fisher's exact test. McNemar's test was used to compare prediction performance of the SVM model and nuclear medicine physician's. Differences in frequencies of scores comprising the probability of PD, either determined by the model or scored by nuclear medicine physicians, were assessed using a χ^2 test. A significance level at 0.05 was used and Bonferroni correction was applied when necessary.

Results

Patient characteristics

Patient characteristics and the various ratios are summarized for PD and non-PD patients in Tables 1, 2, respectively. For PD patients, no differences in characteristics and ratios were found between the training dataset and validation datasets from both hospital 1 ($p > 0.09$) and hospital 2 ($p > 0.16$). Likewise, characteristics and ratios of non-PD patients in group B were similar to group A ($p > 0.2$). No differences were found between non-PD patients in group A and C, except for the $Put_L/Caud_L$ index which was significantly

Table 1 Patient characteristics including age, male, SBRs and putamen/caudate index per center for PD patients

	Group A ($n=58$)	Group B ($n=25$)	Group C ($n=52$)	p value A vs. B	p value A vs. C
Age	68 (61–74)	67 (60–74)	73 (60–78)	0.7	0.2
Male	56.9%	68.0%	69.2%	0.5	0.3
Str. L	2.54 (2.28–2.90)	2.73 (2.41–3.08)	2.67 (2.36–2.95)	0.12	0.2
Str. R	2.52 (2.35–2.71)	2.61 (2.49–2.87)	2.61 (2.34–3.11)	0.14	0.16
Caud. L	3.14 (2.74–3.48)	3.16 (2.95–3.67)	3.32 (2.86–3.68)	0.2	0.16
Caud. R	3.07 (2.81–3.29)	3.16 (2.90–3.52)	3.10 (2.81–3.80)	0.2	0.3
Put. L	2.04 (1.85–2.37)	2.21 (1.93–2.59)	2.04 (1.82–2.45)	0.09	0.9
Put. R	2.05 (1.90–2.25)	2.15 (1.98–2.30)	2.12 (1.82–2.42)	0.3	0.5
Put/caud. L	0.67 (0.62–0.74)	0.66 (0.60–0.77)	0.66 (0.57–0.73)	0.8	0.3
Put/caud. R	0.69 (0.65–0.76)	0.67 (0.63–0.73)	0.68 (0.62–0.75)	0.3	0.5

Data are presented as median (interquartile range) or percentage; *str* striatum; *caud* caudate nucleus; *put* putamen; *L* left; *R* right

The p values are given for either the χ^2 test or Mann–Whitney *U* test

Table 2 Patient characteristics including age, male, SBRs and putamen/caudate index per center for non-PD patients

	Group A ($n=32$)	Group B ($n=15$)	Group C ($n=28$)	p value A vs. B	p value A vs. C
Age	69 (60–78)	71 (66–77)	74 (69–78)	0.7	0.16
Male	37.5%	40%	57.1%	> 0.99	0.2
Str. L	3.87 (3.44–4.36)	3.76 (3.46–4.07)	3.64 (3.23–4.00)	0.5	0.08
Str. R	3.84 (3.31–4.24)	3.78 (3.44–4.01)	3.60 (3.20–3.83)	0.6	0.13
Caud. L	4.31 (3.80–4.95)	3.94 (3.57–4.43)	4.08 (3.73–4.47)	0.2	0.3
Caud. R	4.07 (3.49–4.64)	4.08 (3.50–4.27)	4.04 (3.63–4.32)	0.6	0.3
Put. L	3.60 (3.13–4.13)	3.44 (3.13–4.01)	3.27 (2.58–4.32)	0.8	0.014
Put. R	3.55 (3.07–3.56)	3.51 (3.25–3.75)	3.16 (2.96–3.45)	0.8	0.030
Put/caud. L	0.83 (0.78–0.89)	0.88 (0.83–0.94)	0.76 (0.70–0.82)	0.10	0.002
Put/caud. R	0.87 (0.80–0.90)	0.88 (0.79–0.92)	0.80 (0.73–0.85)	0.7	0.014

The p values are given for either the χ^2 test or Mann–Whitney *U* test

Data are presented as median (interquartile range) or percentage; *str* striatum; *caud* caudate nucleus; *put* putamen; *L* left; *R* right

lower in group C after Bonferroni correction was applied ($p = 0.002$).

SVM classifier development and validation

A regularization parameter value of 2^{-4} was selected to derive the final model, providing a mean F1-score of 0.956 ± 0.002 as determined by the 10-times repeated, stratified tenfold cross-validation. A corresponding mean classification accuracy of 94.3% was found for this parameter value. Validation of the derived model resulted in accuracies of 95.0% and 82.5% for group B and C, respectively, as shown in Fig. 2. The sensitivity, specificity and accuracy of the SVM-model for all three groups are shown in Table 3. Prediction performance of the model for group B was comparable to group C ($p > 0.09$).

Clinical value

Using the combination of visual assessment and ratio interpretation, nuclear medicine physicians were able to discriminate PD from non-PD with an accuracy of 95.0% and 81.3% for group B and C, respectively, as shown in Fig. 2. For both groups, the presentation of ratios in addition to I-123 FP-CIT images to the physician did not provide an increase in accuracy ($p > 0.5$), sensitivity ($p > 0.99$) or specificity ($p > 0.99$). Additionally, scored probabilities of PD as assessed by the physician using only visual assessment was comparable to the probabilities of PD scored using the combination of visual assessment and ratio interpretation ($p > 0.2$), as shown in Fig. 3.

Table 3 Prediction performance of the derived SVM model for all groups

	Group A*	Group B	Group C	*p* value (B vs. C)
F1-score	0.956	0.96	0.85	0.13
Accuracy (%)	94.3	95.0	82.5	0.09
Sensitivity (%)	96.4	96.0	88.5	0.4
Specificity (%)	90.6	93.3	71.4	0.13

Comparing the prediction performance of the SVM model with that of nuclear medicine physicians, similar accuracies ($p > 0.4$), sensitivities ($p > 0.3$) and specificities ($p > 0.99$) were found in both groups when patients assigned to categories 1 and 2 (chance on PD $\leq 49\%$) were classified as non-PD and categories 3 and 4 (chance on PD $\geq 50\%$) were classified as PD", as shown in Fig. 2. In group B, the SVM showed an increase in the confidence of diagnosis as shown in Fig. 3. A higher number of images was scored as 1 (< 20% probability of PD) or 4 (> 80% probability of PD) compared to visual assessment ($p = 0.035$). However, this was not observed in group C ($p = 0.8$). Furthermore, no difference in the frequencies of the scored probability of PD was found between the SVM model and physicians using the combination of visual assessment and ratio interpretation for both centers ($p > 0.8$).

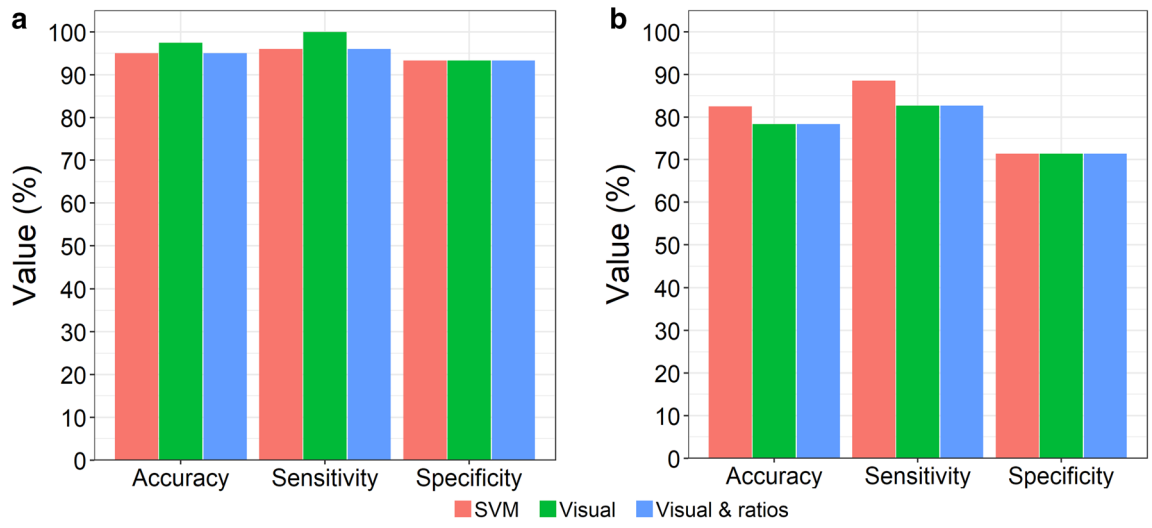

Fig. 2 Column chart showing the prediction performance of visual and semiquantitative interpretation by nuclear medicine physicians and interpretation by the SVM model for the validation set of (**a**) hospital 1 and (**b**) hospital 2. No significant differences were observed between prediction performance variables of the SVM model and nuclear medicine physicians ($p > 0.25$)

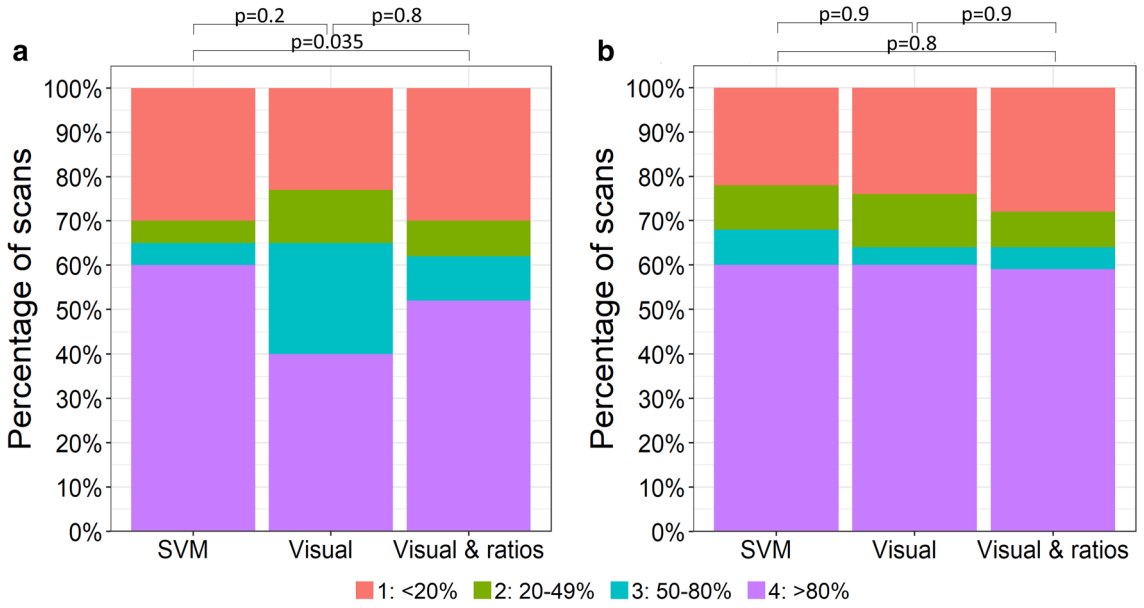

Fig. 3 Column chart showing the percentage of scans for the four different probability-scores for PD as determined by nuclear medicine physicians and the SVM model for the validation set for (**a**) group B from hospital 1 and (**b**) group C from hospital 2

Discussion

In this study, we derived a SVM model to discriminate PD from non-PD with high accuracy using striatal uptake ratios, age and gender as input features. The model was validated using two independent datasets from two different centers. Validation of the model using the unseen dataset from the same center showed a comparable prediction performance, indicating that the model is generalizable towards previously unseen data. For both centers, the performance was similar to that of nuclear medicine physicians who classified patients by visual assessment of I-123 FP-CIT images.

The high classification accuracy of 94.3% for the cross-validation found in this study is in line with several other studies that developed linear SVM models to evaluate SBRs derived from I-123 FP-CIT images [10, 14–16]. Palumbo et al. found an accuracy of 94.2% in a fivefold cross-validation using 90 patients with SBRs and age as input features for discriminating PD from non-PD [14]. Prashanth et al. showed that their model was able to correctly discriminate healthy controls from early PD in 92.3% of cases ($n = 548$) in a tenfold cross-validation using only SBRs [15]. More recently, Taylor et al. compared different ML algorithms with a range of semi-quantification methods and showed that ML generated equal or higher mean accuracies than semi-quantitative analysis, irrespective of the method used. Performing a 10-times repeated tenfold cross-validation, they obtained a mean accuracy of 95% ($n = 657$) and 89% ($n = 304$) for their linear SVM models that were able to discriminate healthy controls from PD and Parkinsonian from non-Parkinsonian patients, respectively [10]. Though high accuracies were found, none of these studies performed a subsequent validation using previously unseen data, nor did they assess whether derived models were generalizable towards other centers.

The prediction performance found for the internal validation, as described above, is not a guarantee for good generalizability as derived models usually perform better on the dataset used for training. Especially in small datasets, a decrease in prediction performance is often seen when the model is tested using an independent unseen dataset. This can be due to overfitting, the lack of representative data used to train the model or the use of different imaging and processing protocols [17, 18]. Usage of the derived SVM model in different centers requires the use of comparable acquisition and image processing methods to ensure that findings are due to underlying pathology and not due to variability in acquisition, reconstruction or quantification settings as these are known to influence striatal uptake ratios irrespective of the presence of a neurodegenerative disease [8, 9, 19]. The similar model prediction found for both group B and C in comparison to group A indicates that overfitting did not occur. Since the different centers employed comparable acquisition and image processing methods, we assume that a decrease in the $Put_L/Caud_L$ index for non-PD patients in group C is due to a difference in patient population.

This study has several limitations. First, a relatively small number ($n = 80$) of patients was used for training purposes that is not necessarily representative of all neurodegenerative and non-degenerative diseases that are

evaluated using I-123 FP-CIT SPECT imaging. Other degenerative conditions that also show reduced tracer uptake are prone to be misclassified by the derived SVM model as I-123 FP-CIT SPECT imaging cannot reliably discriminate between neurodegenerative parkinsonian disorders [19]. Likewise, PD patients with scans without evidence for dopaminergic deficit (SWEDD) will presumably be incorrectly classified as non-PD. Therefore, comparison of the performance of discriminating PD from non-PD by the model and nuclear medicine physicians is needed to put the performance of the model into perspective. The higher number of other neurodegenerative diseases in group C such as patients with drug-induced parkinsonism, essential tremor, multiple system atrophy, Lewy Body dementia and vascular parkinsonism, could explain the lower median values of the ratios of non-PD patients in group A in comparison to group C and thus lower prediction performance of the derived SVM model but also of the physicians as these patients as they are prone to misclassification. To identify neurodegenerative subtypes and SWEDD patients based on I-123 FP-CIT ratios using SVM, a larger patient population is needed with sufficient patients of each subtype. Multiple SVM models can then be derived to discriminate PD from other neurodegenerative diseases as shown by Nicastro et al. [21].

Second, information beyond the uptake ratios combined with gender and age was not considered for input data. As PD is a clinical diagnosis, the addition of parameters comprising the severity and progression of a patient's disease could have allowed better discrimination between PD and non-PD. This would require consistent assessment of patients suspected for PD, thereby using clinical rating scales such as the Movement Disorder Society Unified Parkinson's Disease Rating Scale and Hoehn and Yahr scale [22, 23]. Furthermore, imaging features extracted from I-123 FP-CIT SPECT scans that comprise striatal shape have shown discriminative power in the identification of PD patients [24]. The addition of these features could have provided higher classification accuracies, but requires the development and validation of quantification methods to extract these features. In contrast to this, the current parameters used as input features are routinely collected in clinical practice and therefore easily available. As clinical parameters and visual features of I-123 FP-CIT images can have added value in diagnosing PD, the output of the model is merely a diagnostic aid and should be linked to these clinical parameters by the treating physician for a definitive diagnosis.

Third, different ML approaches including deep learning were not evaluated in discriminating PD from non-PD. Though these approaches could have been superior in performance to that of linear SVM, the derived model is relatively simple and transparent and can be easily exported and used in clinical practice.

Finally, one needs to be aware that the derived model requires a consistent way of assessing the different ratios in order to work properly. Automated approaches could overcome the variability associated with approaches that require manual steps.

Clinical implications

This study provides new insights and has several clinical consequences. The derived SVM model is an objective classification approach for identifying PD patients and has a similar prediction performance as that of standard visual interpretation by expert nuclear medicine physicians. It can therefore facilitate clinical decision-making and diagnosis when used in clinical practice. Taylor et al. evaluated the impact of the addition of SVM-based interpretation of I-123 FP-CIT scans on clinical reporting. They found that consistency between reporters improved and that the model gave added confidence in terms of diagnostic confidence scores. We can therefore assume that usage of our model in clinical practice can lead to less interpretation variation and more confident diagnosis of PD. This model is assumed to be feasible in centers using similar acquisition and image processing methods. For implementation, we advise that each center performs a pilot study to investigate how well the classifier can be generalized to one's own dataset. Ideally, the population characteristics are comparable to that of the training dataset and the model's performance is put into perspective by comparing it with that of nuclear medicine physicians to ensure reliable predictions are generated.

Conclusion

Development of a linear SVM model to interpret I-123 FP-CIT images allows high-accuracy detection of PD with similar classification accuracy as that of expert nuclear medicine physicians. The model is able to discriminate PD from non-PD based on I-123 FP-CIT uptake ratios, age and gender that are collected routinely in clinical practice and are therefore easily available. The model is generalizable towards previously unseen data and feasible to use in centers using comparable acquisition and image processing methods. The results of this study show that the use of the derived SVM model has great potential to be used in the diagnostic process of PD, thereby encouraging implementation of this SVM model in clinical practice.

Acknowledgements None of the authors have anything to disclose.

Compliance with ethical standards

Conflict of interest None.

References

1. Brooks DJ. Imaging approaches to Parkinson disease. J Nucl Med. 2010;51:596–609.
2. Catafau AM, Tolosa E. Impact of dopamine transporter SPECT using123I-Ioflupane on diagnosis and management of patients with clinically uncertain parkinsonian syndromes. Mov Disord. 2004;19:1175–82.
3. Marshall VL, Reininger CB, Marquardt M, Patterson J, Hadley DM, Oertel WH, et al. Parkinson's disease is overdiagnosed clinically at baseline in diagnostically uncertain cases: a 3-year European multicenter study with repeat [123 I]FP-CIT SPECT. Mov Disord. 2009;24:500–8.
4. Bairactaris C, Demakopoulos N, Tripsianis G, Sioka C, Farmakiotis D, Vadikolias K, et al. Impact of dopamine transporter single photon emission computed tomography imaging using I-123 ioflupane on diagnoses of patients with parkinsonian syndromes. J Clin Neurosci. 2009;16:246–52.
5. Djang DSW, Janssen MJR, 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.
6. Darcourt J, Booij J, Tatsch K, Varrone A, Vander Borght T, Kapucu ÖL, et al. EANM procedure guidelines for brain neurotransmission SPECT using 123I-labelled dopamine transporter ligands, version 2. Eur J Nucl Med Mol Imaging. 2010;37:443–50.
7. Booij J, Dubroff J, Pryma D, Yu J, Agarwal R, Lakhani P, et al. Diagnostic performance of the visual reading of 123 I-Ioflupane SPECT images with or without quantification in patients with movement disorders or dementia. J Nucl Med. 2017;58:1821–6.
8. 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 123I-FP-CIT SPECT normal database. Eur J Nucl Med Mol Imaging. 2011;38:1529–40.
9. 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.
10. Taylor JC, Fenner JW. Comparison of machine learning and semi-quantification algorithms for (I123)FP-CIT classification: the beginning of the end for semi-quantification? EJNMMI Phys. 2017;4:29.
11. Taylor JC, Romanowski C, Lorenz E, Lo C, Bandmann O, Fenner J. Computer-aided diagnosis for (123I)FP-CIT imaging: impact on clinical reporting. EJNMMI Res. 2018;8:36.
12. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med. 2019;25:30–6.
13. R Core Team. R: A Language and Environment for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.
14. Palumbo B, Fravolini ML, Buresta T, Pompili F, Forini N, Nigro P, et al. Diagnostic accuracy of Parkinson disease by support vector machine (SVM) analysis of 123I-FP-CIT brain SPECT data. Medicine (Baltimore). 2014;93:e228.
15. 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.
16. Iwabuchi Y, Nakahara T, Kameyama M, Yamada Y, Hashimoto M, Matsusaka Y, et al. Impact of a combination of quantitative indices representing uptake intensity, shape, and asymmetry in DAT SPECT using machine learning: comparison of different volume of interest settings. EJNMMI Res. 2019;9:7.
17. Bleeker S, Moll H, Steyerberg E, Donders AR, Derksen-Lubsen G, Grobbee D, et al. External validation is necessary in prediction research. J Clin Epidemiol. 2003;56:826–32.
18. de Bruijne M. Machine learning approaches in medical image analysis: from detection to diagnosis. Med Image Anal. 2016;33:94–7.
19. Kagi G, Bhatia KP, Tolosa E. The role of DAT-SPECT in movement disorders. J Neurol Neurosurg Psychiatry. 2010;81:5–12.
20. Tossici-bolt L, Dickson JC, Sera T, De NR, Bagnara MC, Jonsson C, et al. Calibration of gamma camera systems for a multicentre European 123 I-FP-CIT SPECT normal database. Eur J Nucl Mol Imaging. 2011;38:1529–40.
21. Nicastro N, Wegrzyk J, Preti MG, Fleury V, Van de Ville D, Garibotto V, et al. Classification of degenerative Parkinsonism subtypes by support-vector-machine analysis and striatal 123I-FP-CIT indices. J Neurol. 2019;266:1771–81.
22. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord. 2008;23:2129–70.
23. Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C, et al. Movement disorder society task force report on the Hoehn and Yahr staging scale: status and recommendations the movement disorder society task force on rating scales for Parkinson's disease. Mov Disord. 2004;19:1020–8.
24. Oliveira FPM, Faria DB, Costa DC, Castelo-Branco M, Tavares JMRS. Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of Parkinson's disease based on [123I]FP-CIT SPECT images. Eur J Nucl Med Mol Imaging. 2018;45:1052–62.