Investigation the use of radiomics for analysis of DAT SPECT imaging in Parkinson's Disease

Alessia Rossi[1], Alessandro Sandron[2], Alesia Memushaj[3] and Valeria Boschetti[4]

University of Padua, Department of Information Engineering.
Biomarkers, Precision Medicine and Drug Development, Prof. M. Veronese, A.A. 2022-23

Abstract—Parkinson's disease (PD) is a prevalent neurodegenerative disorder for which a definitive cure is currently unavailable. The existing treatments primarily aim at managing symptoms associated with the disease. Molecular imaging techniques, such as single-photon emission computed tomography (SPECT), offer valuable insights into the neurochemical alterations in the brain, aiding in the diagnosis and study of neurodegenerative diseases. The objective of this study is to provide evidence supporting the validity of using dopamine transporter (DAT) SPECT imaging for Parkinson's disease. The dataset analyzed comprises 53 subjects, including 33 idiopathic PD patients and 20 healthy controls. Each subject's dataset consists of 178 Radiomics features obtained from DAT SPECT imaging. Through cross-selection analysis, both with and without covariates, a consistent set of features demonstrating a significant difference between patients and controls was identified ($p<0.05$). Subsequently, an assessment of multicollinearity was conducted, leading to the removal of redundant variables. Further reduction of dataset was achieved through a Cross-Validated Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis, thereby retaining the most relevant features. Based on these variables, a combined Radiomics score was constructed, which demonstrated statistical significance in differentiating between patients and controls ($p<0.005$). Various machine learning techniques were applied to classify subjects as healthy or unhealthy using features derived from different stages of the statistical analysis. The results consistently showed promising potential. Furthermore, the association between Radiomics and the severity of clinical symptoms was investigated. The highest correlations were obtained with features based on morphology and histogram intensity. Despite limitations associated with clinical heterogeneity and dataset characteristics, encouraging outcomes were obtained. These findings suggest that with targeted improvements in experimental design, DAT SPECT imaging could serve as a valuable biomarker for PD, enabling improved diagnosis and treatment strategies.

Keywords— Parkinson's disease – DAT SPECT imaging – Radiomics features

I. Background

Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting >1% of the population $\geq$ 65 years of age and with a prevalence set to double by 2030. PD is characterized by tremor, rigidity, slowness of movements and postural instability. As the disease progresses, non-motor symptoms, such as dementia and dysautonomia, may also occur. Cognitive impairment is up to six times more common in individuals with PD than in the healthy population. Specifically, it all starts with the degeneration of dopamine-producing neurons in the brain. This causes striatal dopaminergic deficit and the accumulation of alpha-synuclein in neuronal inclusions. The exact cause of PD is not known, but it is believed to involve a combination of complex genetic susceptibility and environmental factors [1]. Diagnosing Parkinson's disease typically involves a combination of medical history, physical examination, neurological and laboratory tests, and it is challenging to make an accurate diagnosis of PD in the early stages [2]. Single-photon emission computed tomography (SPECT) imaging can be used as a diagnostic quantitative imaging tool for Parkinson's disease, it can also detect the dopaminergic dysfunction in presymptomatic subjects at risk. SPECT scans can detect changes in the dopamine transporter (DAT) levels in the brain, which is reduced in patients with PD. SPECT imaging is usually performed with the use of a radiopharmaceutical agent, which is taken up by the DAT in the brain [3][4]. Generally, the evaluation of DAT SPECT images is conducted via visual inspection, frequently supported by semi-quantitative ratios, such as the striatum uptake ratio (SUR) [4]. Radiomics is a method for extracting high-dimensional quantitative features in images, that results in the conversion of images into mineable data and the subsequent analysis of these data for decision support [5]. Radiomics analysis of DAT SPECT images is expected to be used not only for diagnosis but also for predicting patient prognosis and determining treatment effects as well as better designs of disease modifying trials, with greater power to ascertain efficacy [4] [6].

a. Severity criteria of Parkinson Disease

The related dataset, which will be defined later, includes information from patients with Parkinson's disease. The data are composed of several variables and characteristics that enable the assessment of disease severity and monitor disease progression over time. Some of the widely used scaling in-

dices include the Levodopa Equivalent Daily Dose (LEDD), which provides a measure of the equivalent daily dose of levodopa and other antiparkinsonian drugs taken by patients. There are also assessment tools such as the Unified Parkinson's Disease Rating Scale (UPDRS), which includes specific sections (I, II, III and IV) to assess motor symptoms, activities of daily living, motor complications and motor fluctuations [1]. Other instruments such as the Non-Motor Symptoms Questionnaire (NMSQ) [7], the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) are used to assess patients' non-motor symptoms, cognitive functions and mental status [4]. This dataset provides a comprehensive overview for assessing the severity and management of Parkinson's disease.

b. Research Questions

Starting from a dataset described in the following section of the report, the aim of this work is to investigate whether there are any differences between controls and individuals with PD in Radiomics features or combined scores. Additionally, the study aims to determine if Radiomics features or derived scores are associated with the severity of PD clinical symptoms. The last question investigates whether the Radiomics features are capable to distinguish patients from controls.

As highlighted in previous studies conducted by Li et al. [2] and Ren et al. [5], correlation and redundancy are two prominent aspects of Radiomics features. Taking this into consideration, the initial focus of the analysis was on feature reduction, selecting those features that exhibited a significant difference between patients and controls. The analysis also considered the potential effects of covariates and aimed to capture the most relevant textural information by performing multicollinearity analysis and regression shrinkage analysis.

Subsequently, the study aimed to develop a Radiomics score, following a similar approach as described in the study conducted by Li et al. [2]. The significance of defining a radiomics signature lies in the potential improvement of classification performance when combined with semi-quantitative indicators compared to using semi-quantitative indicators alone, as highlighted in the study by Shiiba et al. [4].

Finally, to ensure that Radiomics features can effectively discriminate between individuals with Parkinson's disease and healthy controls, machine learning algorithms have been applied throughout the different stages of the analysis, aiming to automatically differentiate between the two conditions.

Based on the findings of Rahmim et al. [8] and considering that the feature extraction in the dataset focused on the putamen region, it was anticipated that there would be a significant ability to distinguish and predict patient and control conditions. However, a reduced association between Radiomic features and symptom severity was expected due to the lack of information about the caudate and the absence of clinical measurements for the control group. Taking these factors into account, the association between Radiomics features and clinical conditions was tested, accounting for covariates to mitigate the influence of confounding factors on the results.

Quantitative analysis is of great importance as it may be more sensitive in detecting early stages of the disease and better tracking disease progression. Such efforts align with the goals of initiatives like the Parkinson's Progression Marker Initiative (PPMI) (Parkinson Progression Marker, 2011) to identify biomarkers of PD progression, which is crucial for the development of novel and improved treatments for PD [8].

TABLE 1: RADIOMICS FEATURES

Feature Class	Abbreviation
Local intensity	loc
Intensity-based statistics	stat
Intensity-volume histogram	ivh
Morphology	morph
Intensity histogram	ih
Grey level co-occurrence matrix	cm
Grey level run length matrix	rlm
Grey level size zone matrix	szm
Grey level distance zone matrix	dzm
Neighbourhood grey tone difference matrix	ngt
Neighbouring grey level dependence matrix	ngl

II. MATERIALS AND METHODS

a. DATASET

The dataset was obtain from the PET NODE REPOSITORY available at the King's College London. More specifically, it used a set of Radiomics features collected from 33 idiopathic PD patients and 20 healthy controls matched. The guidelines for the extraction of Radiomics features have been established by the Image Biomarker Standardisation Initiative (IBSI). The IBSI has meticulously defined 11 feature classes, listed in Table 1. By adhering to the IBSI guidelines, the extraction of Radiomics features ensures standardization and consistency across studies, enabling meaningful comparisons and facilitating the identification of robust biomarkers in medical imaging analysis. The used dataset consists of three components: the first one is 'Radiomics', which comprises a total of 177 features. The second component, 'DemographicClinicals', includes demographic characteristics of the examined group of subjects, such as age, gender, education years, height, weight, and body mass index. Relevant medical history, current medical conditions, results of laboratory screens, and any other relevant information will be listed by group. Lastly, the 'Clinical' component contains 9 features and encompasses clinical information about patients.

1. Screening and Clinical/neuropsychological visit

Firstly, a screening visit is conducted to assess the suitability of the subjects. Subsequently, a Clinical/neuropsychological visit encompasses various assessments to evaluate different aspects of the subjects' health. This is useful to provide a comprehensive evaluation of motor and non-motor symptoms, cognitive function, mood, neuropsychiatric symptoms, and quality of life in individuals with Parkinson's disease. The considered components are as follows: General motor function (Movement Disorder Society (MDS) UPDRS part 2 and 3, Hoehn and Yahr (H&Y) staging), non-motor symptoms (MDS-UPDRS part 1; NMSS; SCOPA for autonomic symptoms; PDSS, ESS and RBDQ to detect sleep problems;

MCAS for constipation; PD Fatigue Scale for fatigue, King's Pain Scale for pain; UPSIT for olfactory function), global cognitive function (MMSE, MoCA), mood problems (BDI-II and GDS), anxiety (STAI), apathy (Apathy scale), global neuropsychiatric battery (NPI), Semantic Fluency, Symbol Digit, Benton Judgment of Line Orientation, Hopkins verbal Delayed Recognition False Alarms, Delayed Recognition Hits and Immediate Recall, Letter Number Sequencing and on quality of life (PDQ-39) and exercise (PASE).

2. Main criteria for Inclusion and Exclusion

To be eligible for enrollment in this study, early and levodopa-treated idiopathic PD patients and healthy non-demented (MoCA>25) control subjects must all have the following eligibility criteria:

- **Inclusion and exclusion criteria for all groups:** Firstly, all subjects must have a full understanding of the nature of the study. This step is crucial to ensure that all participants are fully informed and aware of the study details and objectives. Both men and women, aged 30-85 years, are eligible to participate. It is important to note that women of child-bearing potential must have a negative β-hCG test during the screening phase. Men must agree to refrain from donating sperm for the duration of the study and for 3 months after the last administration of SPECT ligands.
- **Inclusion and exclusion criteria for Idiopathic PD group:** Patients included in this group must have idiopathic PD according to the diagnostic criteria established by the MDS Clinical Diagnostic Criteria. It is important to note that these criteria exclude cases of PD induced by drugs or other diseases, as well as individuals who carry PD risk genes such as LRRK2, SNCA, PARK2, and others. The diagnosis of idiopathic PD should have been made after the age of 30 years. Furthermore, patients must be classified between Stage 1 to 3 (inclusive) on the modified HY scale, which serves as an assessment of PD severity. This classification is based on the patient's condition during the ON state. Patients who are drug-naïve, meaning they have never been treated with dopamine agonists or levodopa, are eligible for inclusion. Additionally, patients who are currently on dopamine replacement therapy are also eligible to participate in the study. Instead, are excluded all the patients who had previous surgery for PD and the patients who are treated with duodopa or apomorphine.

3. Imaging assessments

In this study, all participants underwent SPECT imaging using the radioligand [123I]FPCIT, which specifically targets the dopamine transporter system. The SPECT measurements were obtained by quantifying the radioactivity counts in the brain. DAT SPECT images have been acquired approximately 4 hours after injection of [123I]FP-CIT radioligand with an activity of around 185 MBq. The obtained images have been normalized to an [123I]FP-CIT template and the eight most prominent axial slices containing the striatum have been summed. A volume of interest (VOI) template have then been applied to the summed image, and the VOI analyses have been conducted on the putamen both right and left) using the cerebellum region as the reference tissue. Standardized Uptake Value Ratios (SUVRs) have been calculated as the ratio of the putamen VOI count density to the cerebellum count density, which approximates the binding potential. This method has been previously used in studies involving [123I]FP-CIT SPECT. For patients who receive dopaminergic supplementation, the evaluation (including clinical and imaging) has been performed in the OFF state and after a withdrawal period of 12 hours for immediate release and 24 hours for controlled release dopaminergic supplementation.

b. Radiomics features extraction

Features were extracted for each subject using the MIRP Python package (https://github.com/oncoray/mirp). Calculations were performed in 3D: the ROI corresponding to the putamen was resegmented from the SUVR images to eliminate voxels with an intensity value below a threshold equal to 1.2 (selected after visual inspection of the available data). Discretisation was then performed by setting a fixed bin size of 0.0125 for the intensity histogram of the image. This value was set to have around 64 grey levels in the image. When necessary, features were aggregated using the 3D average method.

c. RESEARCH METHODS

1. Investigation on distributions

After a visual inspection of data quality, an assessment was conducted to examine the distribution of data in demographic variables (Demographics) and Radiomic variables (Radiomics) with the aim to ascertain the appropriate tests to employ and select the most suitable statistical analyses. In order to accomplish this, the Lilliefors test, a test of normality, was employed on both patient and control data for each demographic and Radiomic variable. The purpose of this test is to determine whether the data distribution adheres to a normal or Gaussian distribution.

2. Statistical analysis to check if groups are matched

The focus then shifted to determining whether equivalence existed between patients and controls in terms of demographic characteristics, with the aim of ensuring similarity between the two groups. The purpose of assessing equivalence was to minimize the possibility that observed differences between the groups were influenced by factors external to PD. The considered demographic features included age, gender, years of education, height, weight, and body mass index (BMI). As not all variables exhibited a normal distribution according to the normality test, different tests were employed to evaluate equivalence.

For age and height, a t-Test was conducted to compare means since the distribution was normal. Conversely, for years of education, weight, and BMI, a non-parametric test such as the Wilcoxon rank-sum test was utilized. Regarding gender, which is a nominal variable, the Fisher's exact test was employed to assess whether gender proportions differed significantly between the two groups.

3. Cross Sectional Analysis without covariates

The cross-sectional analysis without covariates was performed to compare patients and controls on different Radiomic features and determine if there were significant differences between the two groups. Two different approaches were followed based on the distribution of the data. For the comparison of the two groups, a t-Test was employed when both distributions were Gaussian, while the Wilcoxon-Mann-Whitney rank sum test was used when one or both the distributions was non-Gaussian. Finally, the number of Radiomic features for which the null hypothesis was rejected was calculated. The significance level (α) for the cross-sectional analysis without covariates was set to 0.05. To address the issue of multiple comparisons, a false discovery rate (FDR) test using the Benjamini-Yekutieli method was applied. This test helps control the rate of false positive cases. The corrected p-values were calculated based on a significance threshold of 0.05.

4. Cross Sectional Analysis with covariates

Cross-sectional analysis with covariates is conducted to determine the significance of Radiomics features identified in the cross-sectional analysis without covariates while accounting for the potential influence of covariates and their combined effects. A multiway analysis of variance (ANOVAN) was performed on the Radiomics features identified in the cross-sectional analysis without covariates. Each feature was treated as the dependent variable, the diagnosis as the independent variable, and demographic characteristics such as gender, age, BMI, weight, height, and education years, along with their interaction terms, were included as covariates. The interaction terms represented the interactions between pairs of demographic characteristics. The objective of this analysis was to assess whether the mean responses for the PD and controls groups were significantly different, considering the influence of covariates. The significance level (α) for the cross-sectional analysis with covariates was set to 0.05. Subsequently, a FDR test using the Benjamini-Yekutieli method was applied.

5. Multicollinearity Analysis and Features Selection

To address the issue of multicollinearity arising from the similarity and redundancy of radiomics features, a standardized approach was applied to the input variables and, subsequently, dimensionality reduction techniques were employed to enhance the reliability of the results.

In order to select the most meaningful features, the variance inflation factor (VIF) was computed. The selection process involved two steps: an initial selection was performed independently for each of the 11 feature classes (VIF-within-groups-selection), followed by a second selection among the remaining features from all classes (VIF-between-groups-selection). The selection procedure is composed by the following steps: a linear regression model was constructed using all the chosen features as inputs; for each feature, the VIF was calculated, and variables with a VIF greater than 10 were identified as exhibiting high collinearity with other variables; among these variables, the one with the highest VIF was sequentially removed. This iterative process was repeated until the model was entirely free from collinearity.

To assess the predictive power of each obtained feature in determining the condition (patient or control), a logistic regression analysis was performed. In this analysis, each feature was treated as an independent variable, and the diagnosis served as the dependent variable. Subsequently, a pairwise linear correlation analysis was conducted to examine the relationship between the logistic regression predictions and the actual condition of the subjects. This correlation analysis aimed to assess the accuracy of the logistic regression model in predicting the true diagnostic status. To account for multiple comparisons, FDR correction was applied to the resulting p-values.

As the final step of features selection, a Cross-Validated Least Absolute Shrinkage and Selection Operator (LASSO) Regularization was performed on the reduced data obtained from VIF-based selections. This was achieved using the *lassoglm* function in Matlab. The distribution was set as *binomial* since the response data consists of PD or control status. The Leave-One-Out Cross Validation (LOOCV) method was employed as the validation technique. The convergence threshold for the coordinate descent algorithm was set at 2e-2. Due to the inherent randomness involved in this technique, the same procedure was repeated ten times with different random seeds. The final selected features were those that appeared in the majority of iterations. The purpose of this operation was to mitigate issues related to the dataset size that could potentially impact the results of following analyses to the greatest extent possible.

6. Radiomics Score

By reducing the features set, the aim was to create a robust combined Radiomics score that could provide valuable assistance in accurately discerning patients from controls. This score would serve as a useful tool in facilitating the differentiation and classification of individuals.

The nature of this score is based on model estimates. Therefore, all the procedures explained in the following paragraph were repeated with ten different random seeds to assess the reproducibility and robustness of the results.

Firstly, the dataset reduced from LASSO was divided into training and test sets. Similarly, the response data, which identifies PD vs controls, was partitioned into corresponding training and test sets. A logistic regression model was fitted on the training set using the *glmfit* function in Matlab. The distribution was set as *binomial*, the link function as *logit*, the maximum number of allowed iterations at 1000, and the termination tolerance for the parameters at 2e-2. The coefficient estimates from this model were used to construct the Radiomics score as a linear combination of the selected features from LASSO. In this case, the test dataset was employed to validate the results on a different set of data.

To ascertain the significant difference in the distributions, the score was divided into patients' and controls' scores, and analyzed accordingly. If both distributions followed a normal distribution, a t-Test was performed. Otherwise, the Wilcoxon rank sum test was utilized. The ability of the Radiomics score to differentiate patients from controls was then evaluated using the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve.

Furthermore, a correlation analysis was conducted between the Radiomics score of patients and the severity scales of the clinical symptoms.

7. Classification

This study aimed to classify patients with Parkinson's disease and unaffected individuals using characteristic features. Various machine learning models, including Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and logistic regression (Logit), were employed for the classification analysis.

Initially, all features were utilized to assess their capability in solving the classification task. However, it was observed that all models suffered from overfitting, indicating the need for features selection. As a consequence, both the VIF-based features selections were utilized to enhance the performance of the model. As a final step, the LASSO features extraction was employed with the aim of further improving the model's performance.

To ensure robustness and reduce potential biases, 10 different seeds were tested for each model (except for the initial whole classification). This approach aimed to avoid any dependence of the results on specific computer hardware or software configurations.

To prevent overfitting and enhance generalization performance, a LOOCV strategy was implemented due to the limited dataset size. This strategy involves iteratively training the models on all samples except one, and then evaluating their performance on the excluded sample. This process is repeated for each sample in the dataset. By testing the models on unseen data, leaving out one sample at a time, their true performance can be assessed, mitigating the risk of overfitting to the training data.

In addition to accuracy and the Kappa parameter, the overall performance of the machine learning classifiers was evaluated using the AUC of the ROC curve. The AUC-ROC considers both the sensitivity and specificity of the model, providing a comprehensive assessment of its performance.

8. Association between Radiomics and severity of clinical symptoms

The second research question focused on determining the presence of any association between Radiomic features and the severity of PD clinical symptoms.

Initially, to highlight potential significant correlations between the data, a Spearman correlation analysis was conducted between the Radiomics data of patients and Clinical data.

Secondly, ANOVAN was performed on all the Radiomic features. In each analysis, a different Radiomic feature was selected as the dependent variable, while demographic characteristics (gender, age, BMI, weight, height, and years of education) and their interaction terms were utilized as covariates. The interaction terms were defined as interactions between two of the covariates. The independent variable in each analysis was one of the nine scales representing the severity of PD clinical symptoms explained in paragraph a of background. The purpose of this analysis was to determine if there were significant differences in mean responses for symptom severity, considering the influence of demographic characteristics. The significance level (α) set at 0.05. Consequently, the same Spearman correlation analysis of the previous step was performed using the Radiomics data of the identified significant features.

Additionally, the ANOVAN was repeated by altering the inclusion of covariates. In the first case, no covariates were considered, while in subsequent cases, demographic characteristics were added one at a time. This step allowed for the identification of which demographic characteristic had an impact on the results, specifically for the clinical scale under examination.

Furthermore, to investigate whether there were Radiomic features capable of distinguishing the severity of PD clinical symptoms, a linear regression model was constructed for each individual scale. Only the features that showed significant results in the ANOVAN were included as regressors, along with the demographic characteristics that had demonstrated an influence on the analysis. The symptom severity scales were used as the data to be predicted. To evaluate the acceptability of the predictions, a correlation value was calculated between the predicted values and the actual data. To manage the redundancy of features within the same class, a selection procedure was computed using VIF.

In all these steps, FDR correction was applied to the resulting p-values.

III. RESULTS

First and foremost, the visual inspection of the dataset indicated the absence of missing data and the presence of a Radiomic feature, "ngl_dc_perc", which had a value of 1 fo all subjects. According to the IBSI documentation, this feature may be completely omitted as it evaluates to 1 when complete neighbourhoods are not required, as is the case under IBSI definition. Given that it does not bring any meaningful information to the analysis, it was eliminated from the dataset. The subsequent analysis of the distribution revealed that not all variables exhibit a normal or Gaussian distribution. Consequently, it was necessary to use various statistical tests. Regarding the following phase of statistical analysis, which aimed to assess the matching of the groups, it was determined that patients and controls are indeed matched across all demographic features. This evidence facilitates the acquisition of reliable findings and enables accurate interpretation of them, by effectively differentiating the effects of Parkinson's disease from other potential confounding factors that might impact the outcomes. Below, the results obtained from the various employed methods are presented.

a. Cross-Sectional Analysis

In the cross-sectional analysis without covariates, it was observed that the majority of Radiomics features exhibit significant differences between patients and controls. This difference suggests that the examined Radiomic features are closely associated with the medical condition of the subjects and they can serve as discriminative features for distinguishing and classifying individuals between patients and controls. The FDR analysis revealed that 147 features were deemed significant and retained for further analysis.

Following the cross-sectional analysis with covariates, the number of Radiomics features was reduced from 147 to 143 ($p<0.05$). This reduction suggests that the interaction with covariates weakened the ability of five features to explain the differences between patients and controls. Subsequently, FDR correction was applied, resulting in the identification of only 122 features as true positives.

b. Multicollinearity Analysis and Feature Selection

The VIF-within-groups-selection revealed a significant presence of multicollinearity across almost all of the feature classes, as shown in the first two columns of Table 2. The classes most affected by multicollinearity were Intensity-based statistical features, Intensity histogram features, and Grey level co-occurrence-based features. As a result, the number of features was reduced from 122 to 48 in the first selection, and further reduced to 13 in the VIF-between-groups-selection. Thus, the remaining 13 features effectively summarize the discarded Radiomics features while minimizing the loss of information through the multicollinearity analysis.

Following the logistic regression and pairwise correlation, it was expected to observe positive and statistically significant correlations between the predictions and the diagnosis. The correlation coefficients ranged from 0.45 to 0.81, with p-values lower than 0.0008.

These findings indicate a favorable outcome, as they demonstrate that the selected Radiomic features can significantly distinguish between patients and controls. Moreover, these correlations remained significant even after applying the FDR correction.

In conclusion, the results of the LASSO regularization were consistent across all the different seeds that were utilized. This ultimately led to the final selection of the following 7 Radiomics features:

- intensity kurtosis (stat_kurt),
- intensity-based coefficient of variation (stat_cov),
- asphericity (morph_asphericity),
- centre of mass shift (morph_com),
- large distance low grey level emphasis (dzm_ldlge),
- grey level non-uniformity (dzm_glnu),
- constrast (ngt_constrast).

c. Radiomics Score

As a result of the fitted linear regression model, a Radiomics score was developed for each of the fifteen subjects included in the test set, repeated ten times with different test sets. The calculation involved all coefficient estimates of the model, including the intercept, as shown in the following general formula:

$$RadiomicsScore = \beta 0 + stat\_kurt * \beta 1 + stat\_cov * \beta 2 + morph\_aspericity * \beta 3 + morph\_com * \beta 4 + dzm\_ldlge * \beta 5 + dzm\_glnu * \beta 6 + ngt\_constrast * \beta 7$$

with β as the model fitted coefficients. In different runs, these coefficients had the following patterns: $\beta 0$ was positive, $\beta 1$ was positive, $\beta 2$ was mostly negative, $\beta 3$ was positive, $\beta 4$ was positive, $\beta 5$ was mostly negative, $\beta 6$ was mostly positive, and $\beta 7$ was mostly negative.

In all cases, a significant difference was observed in the distributions of scores between patients and controls, with p-values of the hypothesis tests lower than 0.005. The boxplot depicting the distributions corresponding to the lowest p-value (0.0000003) is presented in Figure 1. Moreover, the resulting AUC values from the ROC curve analysis were consistently high. The minimum value obtained was 0.9259, while the highest value reached 1.

In conclusion, the correlation analysis with the severity scales of clinical symptoms yielded varying correlation coefficients in each run. Some runs resulted in non-significant coefficients, while others obtained high values. The variability of the results poses challenges in their interpretation and the lack of reproducibility causes the inconsistency of this

TABLE 2: MULTICOLLINEARITY ANALYSIS RESULTS: NUMBER OF FEATURES

Feature Class	Start*	VIF within groups Selection	VIF between groups Selection
Local intensity features	2	1	0
Intensity-based statistical features	15	3	2
Intensity-volume histogram features	13	6	2
Morphological features	7	3	2
Intensity histogram features	17	5	1
Grey level co-occurrence based features	18	5	1
Grey level run length based features	13	5	0
Grey level size zone based features	13	6	2
Grey level distance zone based features	9	6	2
Neighbourhood grey tone difference based features	2	2	1
Neighbouring grey level dependence based features	13	6	0

*Number of features before the multicollinearity analysis

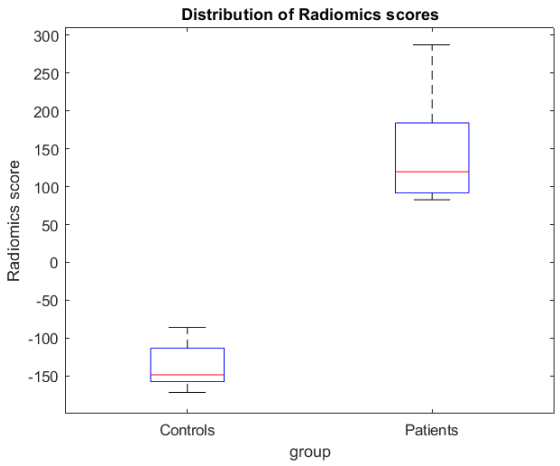

Fig. 1: Boxplot of Radiomics scores of controls vs PD patients

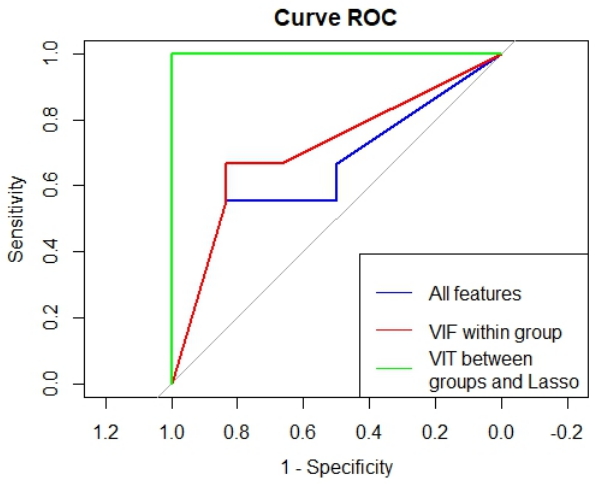

Fig. 2: ROC curves of the most common performance of Logit models for each section

information.

d. Machine Learning Classification

Upon evaluating the individual classification performance of different machine learning algorithms with various runs using different seeds, consistent results were observed.

Initially, the power prediction of all features was examined by running the dataset on SVM, RF, and Logit models. SVM and RF achieved high prediction accuracy, close to 100%, thanks to their complexity and due to the characteristics of the dataset, where the number of features exceeded the number of subjects (not shown in Table 3). However, Logit exhibited lower accuracy of around 60%, with a kappa value of 21%, which can be attributed to the high correlation among the features. The presence of collinearity in the dataset caused instability and implausible values in the estimated coefficients, resulting in decreased power prediction (Figure 2).

Subsequently, polynomial SVM, RF and Logit models were constructed using features from VIF-within-groups-selection. They all showed a slight drop in performance. In the 10 seeds iterations, SVM and RF models consistently maintained high accuracy, while Logit model exhibited a wide range. Its accuracy ranged from 0.3 to 0.93 and kappa value varied between -0.25 and 0.86. The latter finding underscores the fact that the Logit model depends on the choice of seed.

Further, the VIF-between-groups-selection was performed to definitively eliminate collinearity. Polynomial and radial kernel SVM, RF, KNN, and Logit models were trained on this dataset. All of these models achieved correct classifications ranging from 86% to 100% for the patients. They demonstrated high sensitivity (0.83-1) and specificity (0.77-1) in detecting false negative and true negative rates, respectively. Conversely to the previous step, the Logit model exhibited improved performance with an average accuracy exceeding 86% and a kappa value of 73%. This emphasizes the significance of the selected features in effectively classifying the patients.

Finally, the same machine learning models were trained on the dataset extracted from LASSO regularization. In this case, all models exhibited high accuracy and kappa value. Even the Logit model performed well. Indeed, the ROC curve of the Logit models demonstrates a progressive enhancement in its value with each subsequent feature selection step (Figure 2).

TABLE 3: PERFORMANCE OF MACHINE LEARNING ALGORITHMS (RANGES))

Learning Models	Accuracy	Specificity	Sensitivity	Kappa
VIF within group selection				
RF	0.8/1	0.66/1	0.83/1	0.59/1
Pol.SVM	0.73/1	0.55/1	0.83/1	0.5/1
Logit	0.33/0.93	1	0.22/0.88	−0.25/0.86
VIF between groups selection				
RF	0.93/1	1	0.83/1	0.85/1
Pol.SVM	1	1	1	1
Rad.SVM	0,86/1	0.88/1	0.83/1	0.72/1
K-NN	0,86/1	0,77/1	1	0.86/1
Logit	0.86	0.77/1	1	0.73/1
LASSO selection				
RF	1	1	1	1
Pol.SVM	0.93/1	0.88/1	0.83/1	0.86/1
Rad.SVM	0.86/1	0.88/1	1	0.701
K-NN	0.93/1	0.88/1	0.77/11	0.86/1
Logit	0.86/1	0.88/1	0.66/1	0.73/1

e. Association between Radiomics and severity of clinical symptoms

The results of the Spearman correlation analysis, considering all the Radiomics features, revealed significant associations with various scales: UPDRS II, UPDRS III, UPDRS Total, NMSQ, MMSE, and MoCA. MoCA and UPDRS III scales showed respectively the fewest and the highest number of associations, with two and forty-three significant correlation coefficients. In general, the correlation values ranged from 0.34 to 0.53, encompassing both positive and negative values, with p-values lower than 0.049. The highest level of correlation, equal to -0.53 (p-value=0.0014), was observed between the UPDRS III index and the morphological feature "area density (convex hull)". Significant results were

observed across all feature classes, with a majority of morph, ivh and ih features.

The cross-sectional analysis with all demographic characteristics considered as covariates, resulted in the identification of only nine significant features. Subsequent Spearman correlation analysis conducted on these features revealed three significant values, each corresponding to a different scale:

- UPDRS III: a correlation coefficient of -0.48 (p-value = 0.0051) was observed for the morphological "PCA flatness" feature,
- NMSQ: a correlation coefficient of 0.35 (p-value = 0.0475) was found for the intensity-volume histogram "volume at 50% of intensity" feature,
- MMSE: a correlation coefficient of -0.36 (p-value = 0.0370) was obtained for the neighbourhood grey tone difference based "contrast" feature,

as shown in Figure 3.

The analysis on the specific effects of demographic characteristics resulted in the development of individual linear regression models for each of the clinical symptom scales. Among these models, significant correlation coefficients were observed between the actual data and predictions for the UPDRS I, UPDRS IV, and UPDRS Total scales. Specifically, the UPDRS I model obtained a correlation coefficient of 0.36 (p-value=0.0418), the UPDRS IV model obtained a correlation coefficient of 0.42 (p-value=0.0157), and the UPDRS Total model obtained a correlation coefficient of 0.42 (p-value=0.0138). Table 4 provides an overview of the demographic characteristics and Radiomics features used as regressors in these models.

IV. Discussion

In this section, the choices made in the study will be explained and justified, and the results will be discussed.

To control for type I errors resulting from multiple comparison tests conducted in the study, the False Discovery Rate correction method was preferred over the Bonferroni correction. This choice was made because the FDR correction is less conservative and is less likely to discard significant discoveries.

Table 4: Regressors of linear regression models for clinical symptoms scales

Clinical Scale	Demographic Charachteristics*	Radiomics Features**
UPDRS I	A, EY, H	3 ivh
UPDRS IV	G, A, EY H, W, BMI	2 stat, 4 ivh, 5 ih, 1 cm 1 rlm, 1 szm, 1 dzm, 1 ngl
UPDRS Total	G, EY, H W, BMI	1 stat, 2 morph 1 ivh, 1 ih, 1 szm

*A=Age, G=Gender, EY=Education Years, H=Height, W=Weight, BMI=Body Mass Index; **Number for the specific class, abbreviations in Table 1.

Following the cross-sectional analysis, two options were considered: either removing the Radiomics features identified as not significantly important from the model or retaining them along with their dependent covariates. Due to the low number of subjects in the dataset and the high redundancy among the Radiomics features, the first option was chosen to avoid issues related to multicollinearity. This decision aimed to prevent a reduction in statistical power, reliability of estimates, and the complexity of identifying the most important features. To accomplish this, the VIF and LASSO regression analysis were employed.

Selecting the appropriate variable reduction method posed a challenge due to the significant collinearity among the variables. Common techniques in the literature such as LASSO or Ridge regression and Principal Component Analysis (PCA) were unable to yield stable results in this context. Therefore, the calculation of the VIF was considered the safest approach. However, given that the regressor matrix in the linear regression model was close to being singular, an approximation of its inverse matrix was obtained using the *pinv* function in Matlab.

The subsequent features reduction step with LASSO confirms that the multicollinearity analysis and this selection process effectively mitigated redundancy and collinearity issues within the dataset. In fact, the ultimately selected features consistently appeared in each iteration.

These considerations were made evident in the classification procedure. The unreliability of the Logit model's predictions with the dataset from VIF-within-groups-selection can be attributed to the consistent presence of high correlation among the variables. The presence of multicollinearity in the features can result in instability in the estimation of regression coefficients and make it challenging to interpret the individual contribution of each feature to the model predictions. The importance of approaches that reduce redundancy and extract the most relevant features is shown also in the study by Shi et al. [9]. A consistent enhancement in the predictive performance is evident throughout the feature selection procedure, yielding notably favorable outcomes. Notably, the accuracy measure exhibits a prominent increment, particularly in the logit model. Nevertheless, these findings raise concerns regarding potential overfitting of the data or the restricted sample size, as the dataset comprises merely 53 subjects. Despite these considerations, the observed augmentation in accuracy across all models underscores the dataset's capacity to effectively predict the occurrence of the disease. Similar result and approaches can be observed in the study of Iep et al. [10].

Progressing to the Radiomics score construction, through visual inspection of the beta coefficient values calculated using 10 different seeds, it was observed that they exhibit moderate variability. As a result, a single Radiomics score could not be provided. However, it was noted that all 10 formulated scores demonstrate significant ability to differentiate between the patient and control groups. It is presumed that a larger dataset would have yielded more stable and robust estimates of the Radiomics score coefficients across different seeds, thereby enhancing the score's generalizability.

A brief interpretation of the Radiomics score is provided to reinforce the biological significance of Radiomic features in Parkinson's disease.

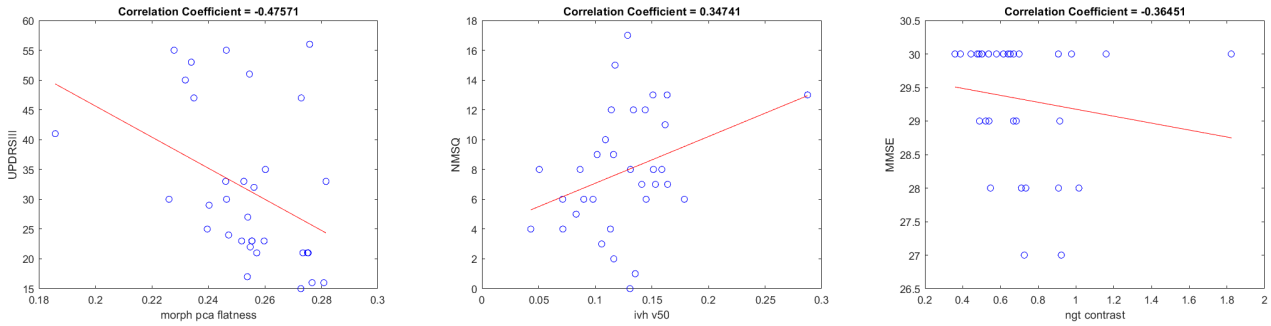

Fig. 3: Correlation plots

In general, it is expected that a DAT SPECT examination of the putamen in a healthy subject will show a region with high biomarker intensity, indicating an abundant presence of functioning dopaminergic receptors. On the other hand, in a subject with Parkinson's disease, a reduction in biomarker intensity in the putamen is expected due to the progressive loss of dopaminergic neurons associated with the disease. Therefore, in the context of radiomic features extracted from DAT SPECT of the putamen, a subject with Parkinson's disease would exhibit a reduction in voxels with high biomarker intensity [3].

The feature "stat_kurt," which indicates the kurtosis of the intensity distribution of voxels, is associated with a consistently positive coefficient, suggesting that in patients, pixel intensities are highly concentrated around a value presumably low (leptokurtic curve), indicating a low accumulation of radiotracer. Similarly, the feature "stat_cov," which describes the dispersion of pixel intensities in the putamen, is associated with a mostly negative coefficient, indicating that in patients, the dispersion is lower and likely confined to low intensities. The feature "ngt_contrast," which describes the intensity variation between voxels and their neighbors, is associated with a mostly negative coefficient, indicating that high contrast is typical in healthy subjects, while patients may exhibit lower contrast and closer, lower intensity levels. Additionally, the feature "dzm_ldgle," which describes the relevance of image regions showing a large intensity difference and low gray level, presents a mostly negative coefficient. This may suggest that the distribution of zones with low gray levels is more uniform in the putamen of Parkinson's patients compared to healthy subjects. The feature "dzm_glnu," which is associated with a mostly positive coefficient, describes the distribution of the number of zones across gray levels. Healthy subjects are characterized by a low value of this feature, indicating a uniform distribution along different gray values, while a higher distribution may indicate the presence of concentrated regions, presumably mostly in lower gray levels, similar to what is described by the features "stat_kurt" and "stat_cov". Regarding the remaining features, their interpretation was more challenging. For example, the feature "morph_asphericity," which is associated with a positive coefficient, indicates that in patients, the deviation of the region of interest (ROI) from a perfect sphere is greater, which could be considered a manifestation of atrophy. However, it should be noted that the putamen can only be approximately considered spherical in healthy subjects, and one must consider that atrophy in the putamen primarily occurs in advanced stages of the disease [11].

A comparison with other studies regarding the selected features for the score would be appropriate only if the image acquisition protocols, reconstruction techniques and correction methods were the same [12]. Nonetheless, a qualitative assessment of the congruence of our results with the literature can be made. For example, the study conducted by Li et al. reports a feature describing contrast in gray levels associated with a negative coefficient and a feature describing the kurtosis of the gray level distribution associated with a positive coefficient, which is consistent with our findings [2]. Similarly, in the study by Shiiba et al., it was demonstrated that putamen features related to uniformity and kurtosis have higher values in patients [4].

Moving on to the analysis of the associations between Radiomic features and the severity of clinical symptoms, an interpretation of the results is presented below.

In the first part of the analysis, morphological features that showed significant coefficients can be interpreted as measures of the complexity or the shape of the ROI, whether it is spherical, compact, elliptical, elongated, or flattened. As discussed earlier for the interpretation of the Radiomics score, these features may be related to tissue atrophy [11]. The results indicate that, for the MoCA and the scales representing motor symptoms (UPDRS II, III and Total), low compactness and low sphericity scores correspond to higher severity symptoms. Regarding the motor symptom scales, features describing how intensities within the ROI are distributed showed correlation values indicating that narrow distributions of grey levels and low intensities are associated with more severe symptoms. As for the features based on the relationship between grey level intensities in the histogram and the volume of the ROI, the scales that have shown significant values pertain to non-motor symptoms and cognitive functions (NMSQ and MMSE). These measures can provide insights into the homogeneity or heterogeneity of intensities within the ROI. The obtained correlation coefficients may suggest that restricted volume portions with the highest intensity values and large portions where the minimum grey level intensity is very low could be associated with greater symptom severity. All values obtained in this first part of the analysis were found to be non-significant after correction with FDR.

To expand the analysis and consider all factors that can influence the various measurements, the ANOVAN was performed taking into account the effects of demographic characteristics and their interactions. Once again, the FDR cor-

rection did not yield significant results. However, a selection of significant p-values was still made, setting a significance threshold of 0.05, given that this is an exploratory study. This allowed for a reassessment of the correlation between only the features considered significant for each scale. The results yielded three correlation coefficients for the UPDRS III scale, the NMSQ scale, and the MMSE scale. Concerning motor symptoms considered in UPDRS III, a negative correlation was obtained with the morphological feature "pca_flatness", which provides a measure of how much the volume of the ROI deviates from a spherical shape. A higher value indicates a more elongated or flattened shape. This finding confirms the previous discussion regarding tissue atrophy. Regarding non-motor symptoms evaluated by NMSQ score, a positive correlation was found with a feature based on the relationship between volume and greyscale intensity "ivh_v50". Specifically, this feature could be interpreted as a measure of intensity dispersion among pixels within the ROI. A low value implies greater variation and heterogeneity of intensities within the region of interest. The positive correlation suggests that greater homogeneity is associated with more severe symptoms. This is in accordance with the behaviour of the findings obtained with the Radiomics score, where low levels of dispersion were associated with patients. Observing the scale assessing cognitive functions MMSE, a negative correlation was obtained with the feature "ngt_contrast" which describe a measure of spatial intensity change. The negative correlation, in this case as well, follows the reasoning used in the Radiomics score and associates low values, hence low intensity variation, with more severe symptoms. The FDR correction in this case confirmed the significance of the first two mentioned correlation values.

In the study conducted by Rahmim at al. [8], correlations indicating an increase in symptom severity with increased heterogeneity of tracer uptake are noted, which seem to contradict the results of the explained analysis. However, a direct comparison would require the same image acquisition and analysis procedure for feature extraction. In the cited study, Haralick features are used, which are mainly based on statistical analyses of image heterogeneity and texture. These can be only compared to the Radiomics features based on the matrix of grey level co-occurrence (cm). Furthermore, in the same study, it is noted that in the analysis with symptom severity, the caudate region showed higher correlations with UPDRS scores. For the MoCA scale instead, it may be more meaningful to consider the concept of laterality, distinguishing between more and less affected areas.

Proceeding with the final part of the conducted analysis, predictions of the linear regression models created for each scale were evaluated against the actual scores of the respective scale. This led to three significant correlation coefficients, although not confirmed by the FDR correction. One involved scale, UPDRS I, evaluates non-motor symptoms of Parkinson's disease. The second involved scale, UPDRS IV, assesses the severity of typical motor symptoms of Parkinson's disease, such as muscle rigidity, tremor, bradykinesia, and posture. The third scale, UPDRS Total, synthesizes the overall evaluation of motor and non-motor symptoms of Parkinson's disease [13]. The analysis led to the results in Table 4, which are not easily interpretable. Some covariates may logically be connected to the severity of specific symptoms, while it is not straightforward to understand how certain factors, height for instance, can influence the scores of non-motor symptoms, such as anxiety, depression, or sleep disturbances.

Nevertheless, this analysis has made it evident that, depending on the scale in object, the specific effects of covariates alone may be more or less significant than the specific effects with the addition of interactions in identifying associations with symptoms severity, highlighting the complexity of this information. As a matter of fact, it is known that environmental factors and personal characteristics of individuals greatly influence these severity scales of symptoms because many of them are linked to cognitive functions or non-motor symptoms related to subjects' lifestyle. The study of Jia et al. [14] for example, has revealed that the level of education significantly influences the performance of the MMSE and MoCA. Isolating these factors simplifies models and may help to understand the role of confounders, leading to greater sensitivity in detecting significant differences between subjects.

Moreover, the study conducted by Chaudhuri et al. [15] notes that non-motor symptoms in PD patients are correlated with disease progression and duration, with higher scores observed in advanced stages. It is important to consider that many factors can change over time, and longitudinal studies would allow for monitoring symptom progression and better interpretation of these scales [6].

In conclusion, from the study by Ren et al. [5], it is evident that significant correlations may not be found with the MMSE and MoCA scales, which lack adequate specificity for the neuropsychiatric features of Parkinson's, which are relevant to the majority of PD patients.

a. Answers to the Research Questions

This study has provided compelling evidence of statistically significant differences between the control group and subjects with PD in the majority of individual radiomics features. Specifically, out of the total of 147 identified features, 122 were found to be independent of covariates and their interactions, effectively distinguishing between the two groups.

Furthermore, it was demonstrated that the created Radiomics score is also capable of distinguishing between patients and controls, offering a valuable diagnostic tool that summarizes the key features. However, due to the limited size of the dataset, it was not possible to establish a unique Radiomics score. A larger sample size would enhance generalizability by capturing a wider range of variations and nuances.

Associations have been identified concerning the interplay between Radiomics features and the severity of clinical symptoms. However, these associations were found to be significantly influenced by the demographic characteristics of the patients and their interactions, thus posing challenges in the interpretation of the findings.

Moreover, based on the previous results of the machine learning algorithms, it can be concluded that the data derived from statistical analysis have the potential to accurately predict the differentiation between healthy individuals and those affected by the disease. However, it is important to acknowledge that due to the size and variability of the data, the ro-

bustness and consistency of the results may be compromised.

b. Limitations and future works

The dataset's size posed practical limitations. In the implementation of LASSO regression, the entire dataset comprising all subjects was utilized without stratifying it into separate training and test sets. Consequently, the chosen features may not be entirely independent of the dataset employed for training machine learning models. Given this issue and the concern about potential overfitting in classification methods, a larger dataset would offer several advantages, including enhanced statistical power, improved generalizability, and more accurate estimation. It is suggested to expand the dataset in further investigations.

To ensure robust and reliable results regarding the association between Radiomics features and clinical measurements, it is suggested to collect clinical indexes for both the control group and the patient group, as it was found to be significant in the study conducted by Rahmim et al. [8]. Furthermore, designing a longitudinal study would be beneficial for this purpose, as demonstrated in the study by Salmanpour et al., where distinct disease trajectories were identified [16]. Similar findings were reported in another study by Rahmim et al. [6]. A longitudinal design would be appropriate also because each subject would act as their own control, helping to normalize the data and compensate for inter-subject confounding factors. It would be interesting to explore the extent to which intra-subject longitudinal analysis improves the applicability of textural features in tracking disease progression at an individual level, as suggested in Rahmim et al.'s study [8].

In the study by Shiiba et al., it was highlighted that lateral differences may be useful for distinguishing between early-stage PD and other forms of Parkinsonism [4]. Therefore, it is hypothesized and suggested that separately analyzing the texture of the left and right putamen could be valuable in distinguishing PD stages of severity.

V. Conclusions

This study highlights the potential of Radiomics analysis in effectively distinguishing patients with PD from healthy controls, underscoring the significance of employing quantitative analysis in PD research. The observed dopamine loss in the putamen, along with the potential value of the caudate in tracking disease progression, supports the decision to focus on the putamen region for feature extraction. However, it should be acknowledged that the limited availability of information about the caudate and clinical measurements for the control group may restrict the association between Radiomic features and symptom severity. Nonetheless, this study contributes to confirm Radiomics analysis for DAT SPECT imaging as biomarker for PD progression, which is essential for the development of improved therapeutic interventions.

a. Software Used

In this study, two software tools were employed for data analysis and model implementation. The following software packages were utilized:

- RStudio (Version: 2022.12.0+353): used to built classification models
- MATLAB (Version: R2023a Update 1 (9.14.0.2239454)): used to perform statistical analysis

b. Other Materials

For more in depth information regarding the way each Radiomics feature is computed, refer to the IBSI documentation (https://arxiv.org/abs/1612.07003)

References

[1] Dag Aarsland, Lucia Batzu, Glenda M. Halliday Gert J. Geurtsen Clive Ballard K. Ray Chaudhuri Daniel Weintraub: *Parkinson disease-associated cognitive impairment*. Nature Reviews Disease Primers, 2021.

[2] Li, Jingwen, Xiaoming Liu, Xinyi Wang, Hanshu Liu, Zhicheng Lin, and Nian Xiong: *Diffusion tensor imaging radiomics for diagnosis of parkinson's disease*. Brain Sciences, 12(7):851, 2022.

[3] Akdemir ÜÖ, Bora Tokçaer A, Atay LÖ: *Dopamine transporter spect imaging in parkinson's disease and parkinsonian disorders.* Turk J Med Sci, 2021.

[4] Shiiba, Takuro, Kazuki Takano, Akihiro Takaki, and Shugo Suwazono: *Dopamine transporter single-photon emission computed tomography-derived radiomics signature for detecting parkinson's disease*. EJNMMI research, 12(1):1–12, 2022.

[5] Ren, Qingguo, Yihua Wang, Shanshan Leng, Xiaomin Nan, Bin Zhang, Xinyan Shuai, Jianyuan Zhang, Xiaona Xia, Ye Li, Yaqiong Ge, *et al.*: *Substantia nigra radiomics feature extraction of parkinson's disease based on magnitude images of susceptibility-weighted imaging*. Frontiers in Neuroscience, 15:646617, 2021.

[6] Rahmim, Arman, Peng Huang, Nikolay Shenkov, Sima Fotouhi, Esmaeil Davoodi-Bojd, Lijun Lu, Zoltan Mari, Hamid Soltanian-Zadeh, and Vesna Sossi: *Improved prediction of outcome in parkinson's disease using radiomics analysis of longitudinal dat spect images*. NeuroImage: Clinical, 16:539–544, 2017.

[7] Vongvaivanich, Kiratikorn, Samart Nidhinandana, Chesda Udommongkol, Parnsiri Chairungsaris, Yotin Chinvarun, Wanna Wongmek, Seema Suphakasem, Jithanorm Suwantamee, and Pasiri Sithinamsuwan: *Non-motor symptoms in thai patients with parkinson's disease studied at phramongkutklao hospital.* Journal of the Medical Association of Thailand= Chotmaihet Thangphaet, 97:S159–67, 2014.

[8] Rahmim, Arman, Yousef Salimpour, Saurabh Jain, Stephan AL Blinder, Ivan S Klyuzhin, Gwenn S Smith, Zoltan Mari, and Vesna Sossi: *Application of texture analysis to dat spect imaging: relationship to clinical assessments*. NeuroImage: Clinical, 12:e1–e9, 2016.

[9] Shi, Dafa, Haoran Zhang, Guangsong Wang, Siyuan Wang, Xiang Yao, Yanfei Li, Qiu Guo, Shuang Zheng,

and Ke Ren: *Machine learning for detecting parkinson's disease by resting-state functional magnetic resonance imaging: A multicenter radiomics analysis*. Frontiers in Aging Neuroscience, 14:806828, 2022.

[10] Iep, Alex, Mohammad B Chawki, Lucas Goldfarb, Loc Nguyen, Vincent Brulon, Claude Comtat, Vincent Lebon, and Florent L Besson: *Relevance of 18f-dopa visual and semi-quantitative pet metrics for the diagnostic of parkinson disease in clinical practice: A machine learning-based inference study*. EJNMMI research, 13(1):1–12, 2023.

[11] Lin, W C, P L Lee, C H Lu, C P Lin, and K H Chou: *Linking stage-specific plasma biomarkers to gray matter atrophy in parkinson disease*. American Journal of Neuroradiology, 42(8):1444–1451, 2021.

[12] Mayerhoefer, Marius E, Andrzej Materka, Georg Langs, Ida Häggström, Piotr Szczypiński, Peter Gibbs, and Gary Cook: *Introduction to radiomics*. Journal of Nuclear Medicine, 61(4):488–495, 2020.

[13] Christopher G Goetz 1, Stanley Fahn, Pablo Martinez Martin Werner Poewe Cristina Sampaio Glenn T Stebbins Matthew B Stern Barbara C Tilley Richard Dodel Bruno Dubois Robert Holloway Joseph Jankovic Jaime Kulisevsky Anthony E Lang Andrew Lees Sue Leurgans Peter A LeWitt David Nyenhuis C Warren Olanow Olivier Rascol Anette Schrag Jeanne A Teresi Jacobus J Van Hilten Nancy LaPelle: *Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan*. pubmed, 2007.

[14] Xiaofang Jia 1, Zhihong Wang 1, Feifei Huang 1 Chang Su 1 Wenwen Du 1 Hongru Jiang 1 Huijun Wang 1 Jiaqi Wang 2 Fangjun Wang 3 Weiwu Su 4 Huifang Xiao 5 Yanxin Wang 6 Bing Zhang 7: *A comparison of the Mini-Mental State Examination (MMSE) with the Montreal Cognitive Assessment (MoCA) for mild cognitive impairment screening in Chinese middle-aged and older population: a cross-sectional study*. pubmed, 2021.

[15] Chaudhuri, Kallol Ray, Pablo Martinez-Martin, Anthony H. V. Schapira, Fabrizio Stocchi, Kapil Sethi, Per Odin, Richard G. Brown, William Koller, Paolo Barone, Graeme MacPhee, Linda Kelly, Martin Rabey, Doug MacMahon, Sue Thomas, William Ondo, David Rye, Alison Forbes, Susanne Tluk, Vandana Dhawan, Annette Bowron, Adrian J. Williams, and Charles W. Olanow: *International multicenter pilot study of the first comprehensive self-completed nonmotor symptoms questionnaire for parkinson's disease: The nmsquest study*. Movement Disorders, 2006.

[16] Salmanpour, Mohammad R, Mojtaba Shamsaei, Ghasem Hajianfar, Hamid Soltanian-Zadeh, and Arman Rahmim: *Longitudinal clustering analysis and prediction of parkinson's disease progression using radiomics and hybrid machine learning*. Quantitative Imaging in Medicine and Surgery, 12(2):906, 2022.