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

Academic Year 2022/2023

Bonato Beatrice
Brusadin Niccolò
Tortelli Mattia
Virgolini Francesca

Abstract:

Parkinson's disease (PD) is a complex neurological disorder that lacks an objective and biologically based gold standard for evaluation, highlighting the need for potential biomarkers. This study investigates the use of radiomics analysis of dopamine transporter (DAT) single-photon emission computed tomography (SPECT) imaging as a potential biomarker for PD. The objectives are to distinguish PD patients from healthy controls (HC) and assess the association between radiomic features and clinical symptom severity.

Three methods of feature selection and six classifiers were implemented to distinguish between PD patients and HC. Among the feature selection methods, two led to promising results: the Lasso and Bottom-Up strategies. The first one identifies a group of radiomics features that perform exceptionally well, with four out of six implemented classifiers achieving perfect accuracy and the remaining obtaining accuracy higher than 90% and sensitivity equal to 1. The latter method also shows promising results, with all classifiers demonstrating 93% or higher accuracy using only two features, more easily derivable if compared to the ones of the Lasso method, therefore offering practical applicability in a clinical environment and potentially speeding up the analysis process. Overall, considering our dataset and taking into account both the variates selection methods and the classifiers implemented, the best combination turns out to be the KNN with the Bottom-Up features (100% accuracy, sensitivity and specificity equal to 1, two features used).

The association between radiomic features and nine PD clinical symptom severity scores was then examined. The same strategies of feature selection previously mentioned were also applied to build several linear regression models, along with the overall correlation between radiomics and evaluation metrics. Results indicate modest outcomes in capturing and predicting clinical severity using radiomics, with some features showing a slight association with PD severity scores. The "UPDRS_I" metric yielded the most effective results with an adjusted R-squared of 0.5311 for Lasso regression and 0.4385 for the Bottom-Up method.

While the overall study's findings provide promising results, several limitations must be acknowledged. The small dataset size hinders the generalizability of the results and the stability of the classification methods. Furthermore, the practical applicability of radiomics in DAT SPECT imaging as a biomarker is limited due to its potential side effects, high cost, and limited availability in healthcare facilities. The existing clinical evaluation methods, albeit subjective, remain easier and quicker to implement. However, the lack of a biologically based gold standard is a crucial and urgent issue to address: the ultimate goal would be to develop biomarkers, also considering radiomics, that can aid in early diagnosis and provide a biologically based evaluation method for Parkinson’s disease. We are optimistic that continuing research in this direction has the potential to shed light on the unknows of this complex field.

Background

Parkinson's disease (PD) is a progressive and degenerative movement disorder characterized by the degeneration of the nigrostriatal dopamine nerve. The evaluation and monitoring of PD traditionally rely on clinical assessments and visual inspection of dopamine transporter (DAT) single-photon emission computed tomography (SPECT) images, which reflect the density of DAT in the striatum. However, these conventional visual approaches have limitations in capturing the full heterogeneity of the disease, as they are primarily based on subjective visual inspection and semi-quantitative ratios.

Radiomics is a technique that extracts and analyzes quantitative features from medical images, providing valuable information about spatial distribution, shape, intensity, and texture patterns within the images. This approach has gained interest in various areas of medical imaging, as it allows for a more comprehensive and objective characterization of diseases. Radiomics has the potential to capture subtle variations in radiopharmaceutical uptake patterns observed in DAT SPECT images of PD patients.

The rationale for applying radiomics to DAT SPECT imaging in PD is multifaceted. Firstly, PD is a heterogeneous disorder with diverse clinical presentations and progression rates, and traditional methods may not capture the complex patterns and subtle changes associated with the disease. Radiomics offers a quantitative and systematic approach to analyze DAT SPECT images, providing a more comprehensive assessment of disease-related alterations. Secondly, radiomics can reveal imaging biomarkers that correlate with important clinical parameters in PD, such as disease severity, motor symptoms, and treatment response. By identifying and quantifying imaging features associated with these clinical variables, radiomics can improve the accuracy of diagnosis, prognosis, and monitoring of PD patients. Additionally, radiomics may uncover novel imaging patterns not easily detected by the human eye, shedding light on underlying disease mechanisms and facilitating the development of targeted therapies. Moreover, the application of radiomics aligns with the trend towards precision medicine, as it can help identify patient subgroups with distinct imaging characteristics or treatment responses, contributing to personalized therapeutic strategies and improved patient outcomes.

The literature provides evidence supporting the use of radiomics analysis in DAT SPECT imaging for PD. Rahmim et al [1] investigated the correlation between Haralick textural features characterizing tracer uptake in DAT SPECT and clinical measures of PD severity and disease duration. Significant correlations between certain textural features and these clinical measures have been demonstrated, indicating the potential of advanced texture metrics as biomarkers of PD progression. Another study by Shiiba et al [2] focused on using a radiomics signature derived from DAT SPECT images to distinguish between healthy controls and PD patients. The researchers found promising results in differentiating between the two groups, suggesting that incorporating radiomics analysis, particularly texture information, can enhance diagnostic accuracy.

Based on the literature evidence, this study aims to further explore the use of radiomics analysis in DAT SPECT imaging for PD. The study has two principal purposes:

1. Investigating group differences between controls and PD patients, examining the differences in radiomic features or combined scores between the two groups and their association with the severity of PD clinical symptoms. We are hypothesizing that significant differences in radiomic features or combined scores will be observed between controls and PD patients, and certain features or scores will be associated with PD clinical symptoms severity.
2. Determining whether radiomic features alone can effectively distinguish PD patients from healthy controls. We are expecting that the radiomic features will exhibit discriminative power for individual prediction.

The study aims to contribute to the growing body of literature on radiomics analysis in DAT SPECT imaging for PD and provide insights into the potential of radiomics as a biomarker for PD diagnosis and assessment of disease severity. The findings may have implications for improving the diagnosis, monitoring, and personalized treatment of PD, leading to better patient outcomes.

Materials and Methods

The data that have been analysed in this report have been derived from the PET Node Repository available at the King's College London. More into detail, it refers to a set of radiomics features of DAT SPECT imaging scans acquired from 33 idiopathic Parkinson Disease's patients and 20 matched healthy controls.

The study from which our dataset is derived focused on assessing and comparing pre-synaptic dopaminergic integrity and function in PD patients and in healthy controls. The involved subjects were preselected after taking into account a list of inclusion and exclusion criteria both in PD group and in HC group. All participants underwent a screening visit, a clinical neurological visit, and an imaging assessment using SPECT with the radioligand [123I] FP-CIT, which targets the dopamine transporter system. Moreover, PD patients on dopaminergic medication were evaluated in both the OFF state (without medication) and after a withdrawal period. Demographic and medical history data were collected and summarized for both groups. SPECT DATSCAN images were obtained approximately 4 hours after injecting [123I] FP-CIT. The images were processed, normalized, and analysed by calculating Standardized Uptake Value Ratios (SUVRs) in the putamen region, using the cerebellum grey matter as a reference tissue. For the radiomics analysis the features were extracted using the MIRP Python package. Discretization was performed on the putamen SUVR images, and features were aggregated using the 3D average method when needed. More detailed information regarding the Node repository can be found at [3].

The resulting dataset, which is the one employed for the analysis in this report, is divided into 3 parts, the largest one containing the radiomics data, while the other two sections contain demographic and clinical information. More specifically, the clinical data is only referred to the diseased subjects and contain some measures related to the severity of the Parkinson's disease. The demographic data instead contains the gender, age education, height, weight and BMI details of the individuals.

All the analysis has been done on R Studio.

First of all, we checked the quality of the given dataset by searching for missing values: the output returned there were none. We then created a new parallel dataset by applying the z-score to the radiomics values. The purpose of this transformation was to use the z-scored dataset in all of the analysis in which differences of variance in the data can significantly impact the results and the original dataset in all the others, thus reducing at the minimum the bias created by the scaling on the data.

The analysis has been divided in two parts: the first one concerned the differentiation between the subjects with Parkinson's disease (PD) and the healthy controls (HC), while the second aimed to find whether there are any radiomic features associated to PD clinical symptoms severity. Here after the procedures employed in both parts will be presented.

1. 'PD VS HC' ANALYSIS

1.1: FEATURE SELECTION

The first step of the first part of our analysis consisted in a feature selection. This step is very important for three main reasons: firstly, dealing with such an high number of features (177) becomes computationally onerous for all the following analysis; secondly, it would be ideal to run an analysis with a concise set of easily computed features rather than relying on a large number of them to obtain good results; last but not least, after examining the radiomics dataset, along with the information contained in the file 'Homework_Radiomics_Manual.pdf', we noticed that many of the features seemed to carry the same type of information, leading us to hypothesize that, due to multicollinearity, many of them could be removed from the analysis to improve its effectiveness.

Two sub-analyses were conducted in parallel. The first focused on feature selection considering only the radiomics data, the second considered both the radiomics data and the demographics' information. The two analyses produced similar results, with slightly better outcomes in the one considering only the radiomics features: therefore, the results here presented will be referred only to the analysis with the smaller dataset.

Initially, the dataset has been divided into training (70% of the individuals) and test set (30%). This was done paying particular attention to the balance of the dataset, since the percentages of the diseased subjects and the healthy controls were not equal. Furthermore, the seed was set to 2. On the training set, three strategies were implemented for feature selection:

- Lasso Logistic Regression: we have given the radiomics features as input to a logistic regressor algorithm, using the Lasso method to shrinkage the parameters. This aggressive method sets to zero all the beta coefficients relative to the features that are already small in the shrinkage-less logistic regression, selecting therefore only the most important covariates. Out of the 177 original features, 11 were selected.

- Bottom-Up strategy: having found the high presence of multicollinearity among the features, we came up with the idea of selecting features by subsequentially adding covariate by covariate to a logistic model. At each step of the iteration, an ANOVA Chi squared test was performed on the created model and only the features with p-value under the 0.05 threshold were maintained for the following iteration. This was done because features with a p-value above this value are probably not statistically relevant in defining the model, since they do not carry further valuable information. With this algorithm, two features were selected.

- Group-by-group strategy: in the 'Homework_Radiomics_Manual.pdf' file, all the radiomics were divided into 11 macro-categories. We decided to divide the radiomics dataset in the same way to examine the relationship among the features belonging to the same macro-group. After testing the correlation (Kendall test, Bonferroni corrected) among them, we found out that variables within the same group were highly correlated. We therefore decided to implement an iterative algorithm for feature selection which builds a model giving as input to a logistic regression model one entire group of features step by step. The algorithm then performs an ANOVA Chi squared test and then selects only the features of the group corresponding to a p-value lower than 0.05 (if there aren't any it selects the feature with the lower p-value). The main purpose of this algorithm is to evaluate the performances taking into account at least one variable for each one of the eleven macro-groups. The final selected features were thirty-three.

1.2: CLASSIFICATION

In order to understand if it is possible to distinguish between subjects with Parkinson's diseased and healthy controls, we implemented six different classifiers. For all of them we performed a tuning of the parameters and a cross-validation on the training set. More into detail, the classifiers and the parameters selected at the end of the tuning procedure were:

- Logistic regression (LR)
- Linear Discriminant Analysis (LDA)
- K-Nearest-Neighbours (KNN): k=9
- Gradient Boosting (GB):
    - Lasso features: n.trees = 5, interaction.depth = 8, shrinkage = 0.09, n.minobsinnode = 2.
    - BottomUp features: n.trees = 5, interaction.depth = 1, shrinkage = 0.1, n.minobsinnode = 5.
    - Group features: n.trees = 5, interaction.depth = 2, shrinkage = 0.1, n.minobsinnode = 2.

- Random Forest (RF): mtry=2.
- Support Vector Machine (SVM): cost=0.1, kernel=linear.

For all the classification methods and the three features selection strategies, the Receiver Operating Characteristic (ROC) was computed, alongside with the Area Under the Curve (AUC), sensitivity and specificity. This was mainly done to compare the classifiers' AUCs, since a higher AUC is associated with a better classifier performance in distinguishing between the two classes.

Looking at the obtained results, we noticed that the Gradient Boosting classifier did not reach the same high performances of the other classifiers, so we decided that it was not worth considering it for further analysis. The results for this classification have however been reported in the relative results tables.

1.3: STATISTICAL ANALYSIS

In order to evaluate the statistical differences between the two groups, we decided to conduct several statistical tests:

- Shapiro-Wilk test: to assess which statistical tests to use in the following analysis, we performed initially a Shapiro-Wilk test on all the features to verify their normality. The null hypothesis for the Shapiro-Wilk test is that the data follow a normal distribution. The test calculates a statistic, W, which is based on the covariance between the sample values and the expected values under the assumption of normality. Of the 177 given features, 146 rejected the null hypothesis, thus showing a non-gaussian distribution. In order to better visualise the obtained results, the histograms and the Q-Q plots of all features' distribution were plotted.

- Wilcoxon Rank Sum test on the single features: since for many of the features the statistical distribution could not be assumed gaussian we performed a Wilcoxon Rank sum test to assess the statistical differences between the two groups. This was done considering one by one all the features: the results are shown via a boxplot.

- De-Long Test: this test verifies if there are any statistically significant differences between the ROC curves of two different models. The analysis was performed on all the previous classifiers (except for Gradient Boosting) and on a logistic regression classifier that considered all the radiomics features. Comparisons were done one by one, together with the computation of the final hypothesis and the corresponding p-value. All the ROC curves were then plotted against the ROC curve of the logistic regression classifier that had all the features as input.

2. PARKINSON'S DISEASE SEVERITY ANALYSIS

For this part of the work, the clinical information contained in the 'CLINICAL' sheet was taken into consideration. The file contains nine columns of metrics data that aimed to assess the severity of the Parkinson's disease. More specifically:

- LEDD TOTAL (L-dopa equivalent daily dose): it provides an artificial summary of the total daily medication a patient is receiving.

- UPDRS (Unified Parkinson's Disease Rating Scale): it consists of an empirical measure of the Parkinson's Disease severity. It is obtained through a test divided into four parts, where four diseases-related aspects (non-motor experiences of daily living, motor experiences of daily living,

motor examination, motor complications) are evaluated via a series of questions. For each one of them, the subject has to give a score that goes from 0 to 5. A lower score represents a normal situation, while a higher one indicates the presence of some complications.

- NMSQ (Non-Motor Symptoms Questionnaire): it is another empirical measure of the Parkinson's Disease severity. Several questions are presented to the subjects related to the presence of Parkinson's symptoms during the previous month. The number of subject's Parkinson's symptoms are then counted to obtain the final NMSQ score.

- MMSE (Mini Mental State Examination): it is a neuropsychological test that aims to evaluate the presence of subjects' cognitive impairment. It consists of a series of questions belonging to seven cognitively different tasks areas and the subject can obtain a score that goes from 0 to 30. If the score is lower than 18, there is a severe impairment of the cognitive abilities, from 18 to 25 a moderate impairment, while scores of 26 and above are associated with normal cognitive abilities.

- MoCA (Montreal Cognitive Assessment): it is another test which aims to evaluate the presence of subjects' cognitive impairment. It is a 10-minute evaluation with a series of questions: the score ranges from 0 to 30. A subject is considered cognitively normal if they obtain a score higher than 26.

2.1: CORRELATION ANALYSIS

The aim of this part of the work is to investigate whether there are any radiomic features that can be put into relationship with the disease's severity. Only the radiomic features belonging to Parkinson's disease patients were considered. In addition, we decided to also take into account the demographic information relative to the age, education, height, weight and BMI of the subjects, to evaluate their influence on the severity assessment. The gender factor was not considered due to its categorical values. The current dataset was then divided into training and test set, with the usual 0.7/0.3 ratio.

We firstly examined the correlation among the given severity metrics: the assumption was that they would exhibit a high correlation if a subject that has a severe form of the disease for one of these metrics has also a severe form of disease for the others. Moreover, we hypothesize that a more severe disease might be associated with a higher drug intake and a more serious cognitive impairment. These hypotheses reveal themselves to be pseudo-correct, since the correlation values are lower than we expected: this discrepancy may be attributed to the fact that the metrics derive from an empirical analysis and not an objective measure.

Despite this finding, we decided to check if there were any correlation among all the radiomics features, the demographic variates and all the considered metrics: this was also done because even if the metrics themselves are not highly correlated with each other, it is still possible that some radiomic features are correlated with one or more of the metrics. We therefore computed the Kendall's correlation and the corresponding p-values (Bonferroni corrected) for all the data previously mentioned. We set then a threshold of 0.4 to highlight the most correlated pairs. The results obtained were stored for the following analyses.

2.2: FEATURE SELECTION

In order to understand if the metrics values could be predicted, we decided to fit a linear regression model with the radiomics features. Since 182 values (5 demographic variates in addition to the 177 radiomics features) are too many to consider, we firstly performed a feature selection, following the same scheme employed in the 'PD vs HC' analysis:

- Lasso Linear Regression: for each one of the 9 metrics above mentioned, a linear regression algorithm has been developed, using the Lasso method to shrinkage the radiomics parameters of the training set.

- Bottom-Up strategy: the same algorithm described in the first part of the analysis has been implemented for the training set, for all the nine different severity metrics.

2.3: LINEAR REGRESSION ANALYSIS

For all the linear regressors described in the previous section we computed the Residual Mean Squared Error (RMSE), the R squared coefficient (R2), the adjusted R squared coefficient (R2 adjusted) and the Akaike Information Criterion (AIC).

We then developed two additional linear regressors to assess the performances obtained considering the results of the correlation analysis, which highlighted a higher correlation among some of the radiomics and the UPDRS I and MMSE metric. The before mentioned performance parameters were also computed for these regressors.

Results

In the present section the results of our analyses will be presented.

1. 'PD VS HC' ANALYSIS

1.1: FEATURE SELECTION

In the first table (table 1) the features selected with the three different methods are shown. It is possible to notice some overlaps between the features selected with the different strategies, although not many: specifically, 'stat_mean' has been selected both for the Bottom-Up and the Group-by-Group methods and four out of the eleven Lasso features ('stat_skew', 'morph_comp_1', 'morph_sphericity', 'ih_skew_fbs_w0.0125') are also included in the Group covariates.

TABLE 1:

LASSO FEATURES	stat_skew, stat_qcod, ivh_diff_v25_v75, ivh_diff_i25_i75, morph_comp_1, morph_sphericity, morph_area_dans_aee, ih_skew_fbs_w0.0125, ih_iqr_fbs_w0.0125, ih_max_grad_fbs_w0.0125, ngl_dcnu_norm_d1_a0.0_3d_fbs_w0.0125
BOTTOM UP FEATURES	stat_mean, stat_var
GROUP FEATURES	loc_peak_glob, stat_mean, stat_skew, ivh_v10, ivh_v25, ivh_v50, ivh_v75, morph_volume, morph_area_mesh, morph_comp_1, morph_sphericity,, morph_pca_maj_axis, morph_pca_min_axis, morph_pca_least_axis, morph_pca_elongation, ih_mean_fbs_w0.0125, ih_skew_fbs_w0.0125, cm_joint_max_d1_3d_avg_fbs_w0.0125, cm_joint_avg_d1_3d_avg_fbs_w0.0125, cm_diff_avg_d1_3d_avg_fbs_w0.0125, rlm_sre_3d_avg_fbs_w0.0125, rlm_hgre_3d_avg_fbs_w0.0125, szm_sze_3d_fbs_w0.0125, szm_lze_3d_fbs_w0.0125, szm_hgze_3d_fbs_w0.0125, dzm_sde_3d_fbs_w0.0125, dzm_lgze_3d_fbs_w0.0125, dzm_sdlge_3d_fbs_w0.0125, ngt_contrast_3d_fbs_w0.0125, ngt_busyness_3d_fbs_w0.0125, ngt_complexity_3d_fbs_w0.0125, ngl_lde_d1_a0.0_3d_fbs_w0.0125, ngl_hde_d1_a0.0_3d_fbs_w0.0125

Table 1 shows the features selected with the three methods. The features common to the Lasso and the Group-by-Group strategy are highlighted in yellow, while the ones present both in the Bottom-Up and the Group-by-Group are highlighted in light blue.

1.2: CLASSIFICATION

The classification results are shown in the table 2. Overall, all classifiers performed impressively well with almost all the selected methods. More specifically, four classifiers (Logistic regression, KNN, SVM, LDA) achieved a balanced accuracy of 100% (Sensitivity=1, Specificity=1, AUC=1) when dealing with the Lasso features, meaning that they were able to discriminate perfectly among healthy controls and subjects with Parkinson's Disease. The same result has been achieved also for what concerns the Bottom-Up features by the KNN classifier and by the Random Forest and SVM algorithms for the Group-by-Group strategy.

TABLE 2:	LASSO FEATURES	BOTTOM-UP FEATURES	GROUP FEATURES
LOGISTIC REGRESSION	Accuracy: 100% Balanced accuracy: 100% SE: 1 SP: 1 AUC: 1	Accuracy: 93.3% Balanced accuracy: 91.67% SE: 1 SP: 0.8333 AUC: 0.9167	Accuracy: 53.33% Balanced accuracy: 52.78% SE: 0.5556 SP: 0.5 AUC: 0.5278
K-NEAREST NEIGHBOURS	Accuracy: 100% Balanced accuracy: 100% SE: 1 SP: 1 AUC: 1	Accuracy: 100% Balanced accuracy: 100% SE: 1 SP: 1 AUC: 1	Accuracy: 93.33% Balanced accuracy: 91.67% SE: 1 SP: 0.8333 AUC: 0.9167
GRADIENT BOOSTING	Accuracy: 93.33% Balanced accuracy: 91.67% SE: 1 SP: 0.8333 AUC: 0.9167	Accuracy: 93.33% Balanced accuracy: 91.67% SE: 1 SP: 0.8333 AUC: 0.9167	Accuracy: 93.33% Balanced accuracy: 91.67% SE: 1 SP: 0.8333 AUC: 0.9444
RANDOM FOREST	Accuracy: 93.33% Balanced accuracy: 91.67% SE: 1 SP: 0.8333 AUC: 0.9167	Accuracy: 93.33% Balanced accuracy: 91.67% SE: 1 SP: 0.8333 AUC: 0.9167	Accuracy: 100% Balanced accuracy: 100% SE: 1 SP: 1 AUC: 1
SUPPORT VECTOR MACHINE	Accuracy: 100% Balanced accuracy: 100% SE: 1 SP: 1 AUC: 1	Accuracy: 93.33% Balanced accuracy: 91.67% SE: 1 SP: 0.8333 AUC: 0.9167	Accuracy: 100% Balanced accuracy: 100% SE: 1 SP: 1 AUC: 1
LINEAR DISCRIMINANT ANALYSIS	Accuracy: 100% Balanced accuracy: 100% SE: 1 SP: 1 AUC: 1	Accuracy: 93.33% Balanced accuracy: 94.44% SE: 0.8889 SP: 1 AUC: 0.9444	Accuracy: 73.33% Balanced accuracy: 72.22% SE: 0.7778 SP: 0.6667 AUC: 0.7222

Table 2 shows the results obtained for all the classification methods and for all the three strategies of feature selection. Dark green fields represent the combination of selected features/classifiers that reached perfect accuracy; light green ones combinations with an accuracy higher than 90%; yellow squares indicate classifiers with performances above 60%, while red ones performances under 60% of accuracy.

The remaining combinations of selected features/classifiers still provided remarkable results (balanced accuracy higher than 90%, sensitivity=1), while having a lower specificity (0.83 in each one): this was the case of the combinations Lasso/GB, Lasso/RF, BottomUp/LR, BottomUp/GB, BottomUp/RF, BottomUp/SVM, Group/KNN and Group/GB. This is indeed a valid outcome, since having sensitivity equal

to 1 means having minimized the false negative cases and thus having correctly classified all the subjects that truly have the disease, despite maybe misclassifying some patients that are negative but are considered positive: these cases can however be managed by doing further tests to validate the previous findings. Regarding the combination BottomUp/LDA, it has given an accuracy higher than 90%, but specificity equal to 1 and 0.8889 sensitivity. This could be a strategy employed in studies in which maximizing the specificity is important, also at the cost of having a lower sensitivity.

It is important to highlight the presence of two outcomes that are far worse than the others previously discussed, both involving the prediction based on the Group-by-Group features. The classifiers at issue are the LDA (balanced accuracy=72.22%, sensitivity=0.7778 and specificity=0.6667) and the Logistic regressor (balanced accuracy=52.78%, sensitivity=0.5556, specificity=0.5). Concerning the LDA, the poor outcomes could be associated with the fact that this type of classifier performs well when the data show a gaussian distribution: as we will later show, not all the considered features satisfy this assumption. We thus hypothesize that the high number of them identified by the Group feature selection process could lead to a higher error rate and worse classification for this method. Regarding the logistic regression, it is well known to perform badly in the presence of a high number of covariates, as in the case of this study.

1.3: STATISTICAL ANALYSIS

As previously mentioned in the materials and methods section, the output of the Shapiro-Wilk test highlighted the presence of many features that do not present a normal distribution, validating the considerations done for the poor results obtained by the LDA classifier. More specifically, only 31 features out of 177 did not reject the null hypothesis.

The Wilcoxon rank-sum test ran on all the features of the dataset returned eighteen accepted hypotheses (meaning that for these variates there were no statistically significant difference between the medians among the PD and HC subjects) and 159 rejected hypotheses (meaning that indeed for these radiomics there were a difference for the two groups). The names of the features that resulted in accepting the H0 hypothesis are shown in the table 3. Notably, some of these belong to the Group and the Lasso selected features, respectively highlighted in yellow and green in the table. This is a peculiar finding, since we hoped to find out also statistically significant differences for all the selected features in the statistical analysis: however, not necessarily the lack of differences in the median result in a poor classification, as proven by our previous results. It is also worth to say that the features selection methods, in particular the group-by-group one which picked many covariates, could also return features that, taken into account individually, are not statistically powerful,

TABLE 3:

FEATURES – H0 accepted
morph_area_mesh
morph_diam
morph_pca_maj_axis
morph_pca_least_axis
morph_pca_elongation
morph_pca_flatness
morph_area_dens_aabb
morph_area_dens_aee
morph_area_dens_conv_hull
morph_geary_c
ih_min_fbs_w0.0125
ih_max_grad_fbs_w0.0125
cm_inv_diff_norm_d1_3d_avg_fbs_w0.0125
cm_inv_diff_mom_norm_d1_3d_avg_fbs_w0.0125
cm_corr_d1_3d_avg_fbs_w0.0125
dzm_lde_3d_fbs_w0.0125
dzm_zd_var_3d_fbs_w0.0125
ngt_coarseness_3d_fbs_w0.0125

Table 3 shows the names of the features for which the Wilcoxon rank-sum test accepted the null hypothesis. Features belonging to the Group-by-Group and to the Lasso selected features are highlighted respectively in yellow and in green.

but on the whole could add value to the feature selection model, as shown by the Anova Chi-squared test performed.

However, the vast majority of the features selected show also a significant difference in the median, which tells us that the features' distributions are statistically different between HC and PD subjects. Figure1.a and Figure 1.b show the boxplot results obtained for the Wilcoxon rank-sum test, respectively for the features belonging to the Lasso and the Bottom-Up groups. Due to their dimensionality, Group-features boxplots (and all the other features' distributions) have been computed but are not showed in this report.

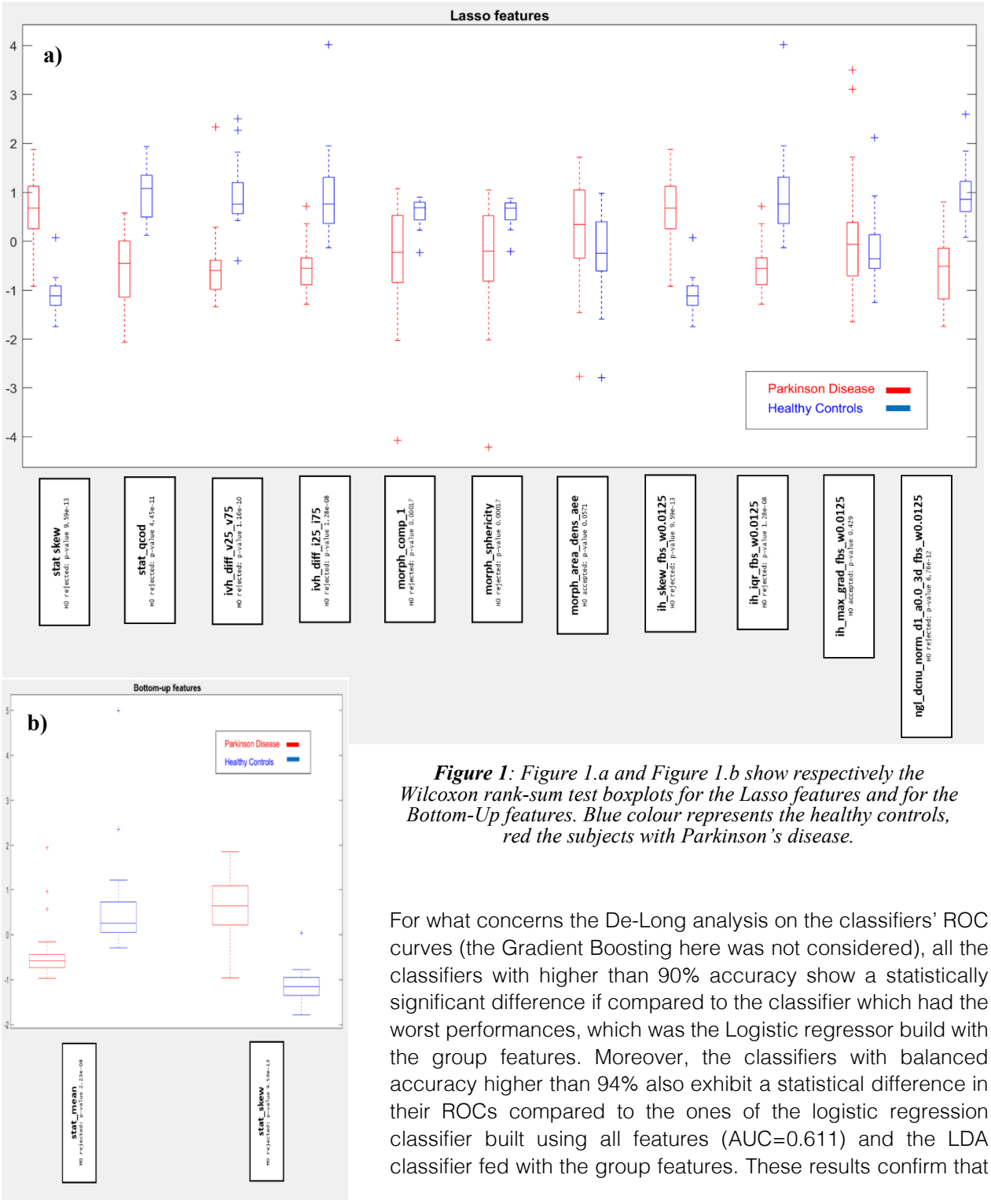


***Figure 1**: Figure 1.a and Figure 1.b show respectively the Wilcoxon rank-sum test boxplots for the Lasso features and for the Bottom-Up features. Blue colour represents the healthy controls, red the subjects with Parkinson's disease.*

For what concerns the De-Long analysis on the classifiers' ROC curves (the Gradient Boosting here was not considered), all the classifiers with higher than 90% accuracy show a statistically significant difference if compared to the classifier which had the worst performances, which was the Logistic regressor build with the group features. Moreover, the classifiers with balanced accuracy higher than 94% also exhibit a statistical difference in their ROCs compared to the ones of the logistic regression classifier built using all features (AUC=0.611) and the LDA classifier fed with the group features. These results confirm that

the classifiers that show perfect accuracy have significantly improved performances if compared with the majority of all other classifiers and thus outperform them in terms of discriminatory power between the PD and HC groups. The ROC curves and their AUCs can be better visualized in the figure 2.

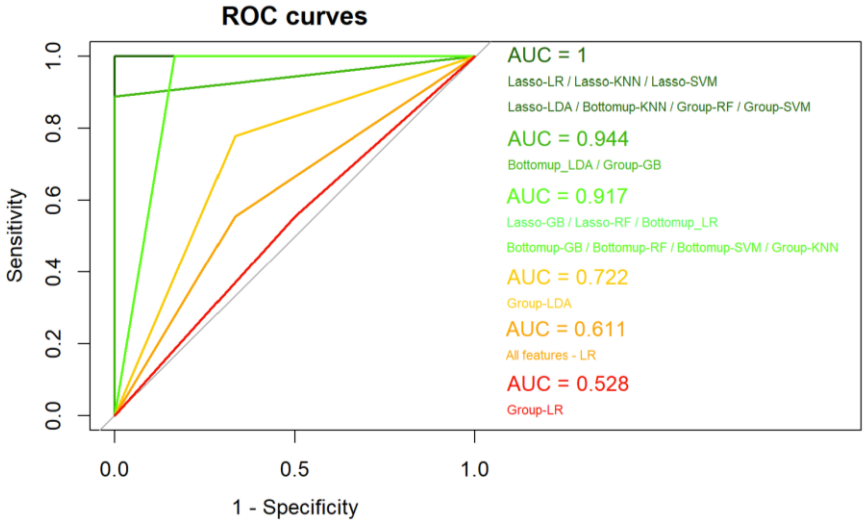


Figure 2: *The ROC curves and the AUCs of every implemented classifier. The colours range from dark green to yellow to red according to the decrease in the performances of the considered classifiers.*

2. PARKINSON'S DISEASE SEVERITY ANALYSIS

2.1: CORRELATION ANALYSIS

Figure 3.a shows the heatmap of the Kendall correlation matrix among all the radiomic features, the demographic variates and the evaluation metrics of the Parkinson's disease. The presence of numerous squares along the main diagonal indicates the existence of multicollinearity among the features within the same macro-group, supporting our hypothesis. The dark red little square at the bottom right extremity of the matrix image represents the Kendall correlation values among the evaluation metrics described in the materials and methods section. These values are better shown in Figure 3.b.

From Figure 3.b it is possible to see that the MoCA and the MMSE values are the metrics which show a negative value of correlation with the other variables, correlation that appears to be relatively lower in magnitude if compared to the correlations among the remaining metrics. This is exactly what we expected: firstly, the MoCA and MMSE metrics represent a sort of measure of the subjects' cognitive impairment, whereas the other metrics asses the severity of the Parkinson's symptoms or the drug intake of the subjects, consequently justifying the presence of lower absolute correlation values. Secondly, the MoCA and MMSE ranges are inverted if compared to the other metrics' ones, meaning that a lower value in these two variables is associated with healthier patients: this does not happen with the other tests' results in which by increasing the score also the degree of Parkinson severity increases. In this way, the negative correlation values, particularly highlighted for the NMSQ comparisons, are in agreement with what we were expecting.

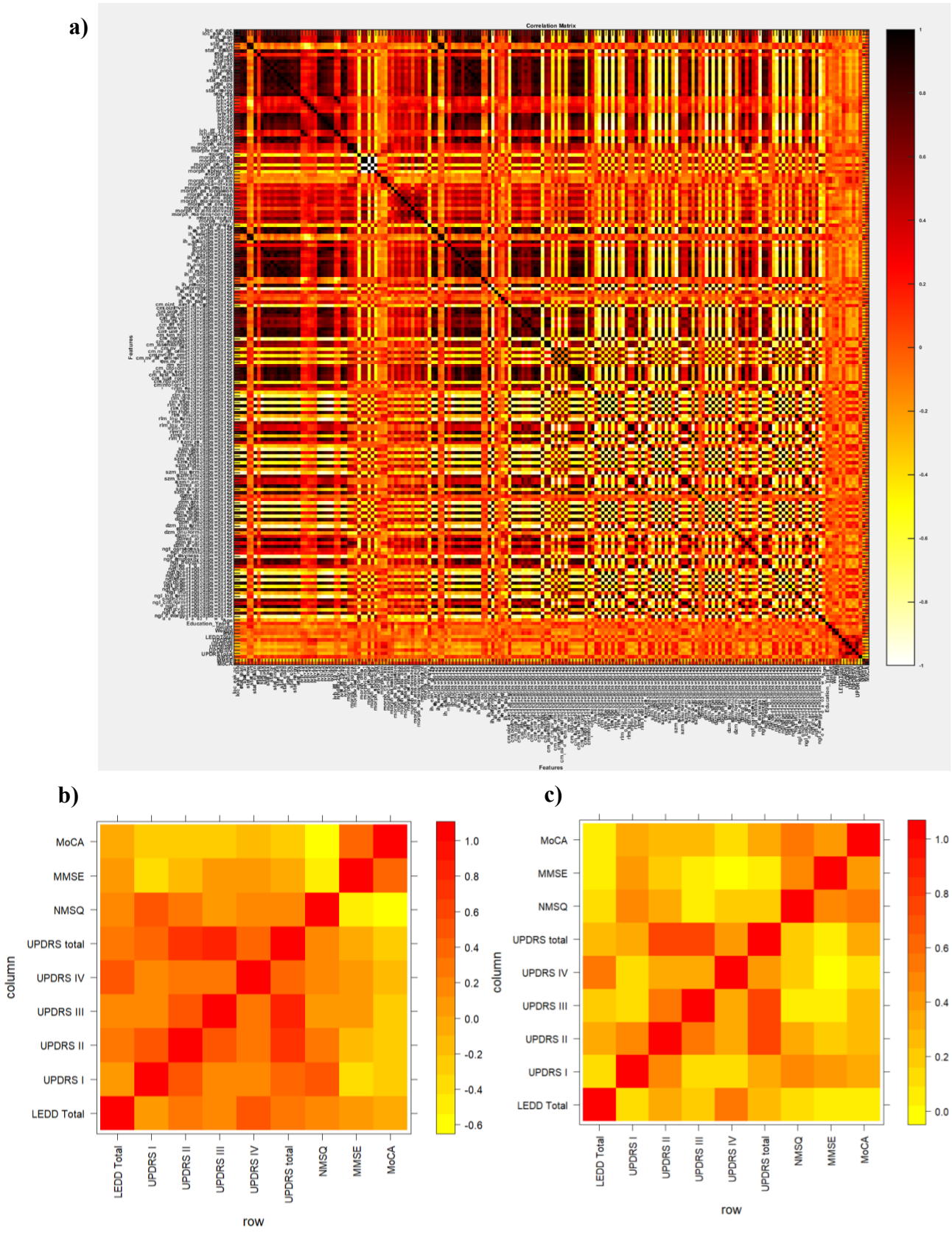


Figure 3: *a) shows the heatmap of the Kendall correlations among all the radiomics, the demographic information and the evaluation metrics; b) highlights the correlation values among the metrics themselves; c) shows the same correlation matrix as b) but on which the absolute values have been computed.*

Figure 3.c provides a visualization of the absolute correlation values, emphasizing the correlations that significantly deviate from zero. The highest correlation values are observed between UPDRS Total vs UPDRS III and UPDRS II: this was predictable since UPDRS Total combines the UPDRS scores altogether.

Reconsidering figure 3.a, it is crucial noting that the correlation among the clinical metrics and the radiomic features is far lower than the correlation among the features themselves, proven by the fact that the image shows two mainly lighter coloured red symmetrical bands where these comparisons are computed (however, part of the red bands is also due to the poor correlation values between the demographic information and the radiomic values). This justifies the 0.4 absolute correlation threshold we decided to set to select the most correlated features in relation to the various metrics. We found that UPDRS I is well correlated (where by 'well' we mean 'above the set threshold') with 'ivh_v90' (r=0.4139, p-value=0.0067), while MMSE is well in relationship with 'stat_rmad', 'stat_qcod', 'ivh_v90' and 'ih_rmad_fbs_w0.0125' (absolute r values higher than 0.404, p-values lower than 0.02). These values are shown more in detail in table 4.

From the literature, we are aware that MMSE scores higher than 26 indicate normal cognitive functioning. Upon visual inspection of the data, it appears that none of the subjects have scores lower than this threshold, suggesting that all subjects in our dataset are cognitively normal according to the MMSE test. This was not the case for the MoCA scale, that highlighted 8 abnormal subjects. The results in output of the correlation are therefore particularly surprising since we expected to find higher correlation with the MoCA values instead of with this metric. However, although this is a useful result to start our analysis with, the presence of correlation with the MMSE scale does not necessarily mean that there is a biological relevance underneath. Moreover, among the features selected by the correlation there could be some that carry the same information: this seems to be the case of 'stat_rmad' and 'ih_rmad_fbs_w0.0125', which have the same correlation and p-value against MMSE and in the overall Kendall matrix are correlated among themselves with an approximated value of 1, proving therefore our point.

2.2: FEATURE SELECTION AND LINEAR REGRESSION ANALYSIS

Onwards with the analysis, we considered the same training set used in the correlation analysis to investigate any potential associations between certain radiomics features and the severity of PD clinical symptoms.

As previously mentioned in the materials and methods section, we implemented Lasso linear regression on the PD clinical scores, only getting results for the following variables: "LEDD_TOTAL," "UPDRS_I," "UPDRS_III", and "MMSE." One of the reasons why Lasso Regression did not yield any features among all the radiomics for certain dependent variables (resulting in all beta coefficients being zero) could be because these variables are clinical scores assessed in a rather subjective manner. Clinical scores, as in our case, often involve human judgment and interpretation, which can introduce variability and subjectivity in the measurements. As a result, it might become challenging for Lasso Regression to identify specific features that consistently explain the dependent variable among the radiomics. As opposite to the clinical metrics, the independent variables (radiomics) are quantitative variables derived from medical imaging data. These radiomic features provide objective and quantifiable information, making them more amenable to analysis using Lasso Regression. Moreover, Lasso algorithm is known for its ability to perform feature selection and regularization by shrinking coefficients towards zero. When faced with a set of independent variables like radiomics that are quantitative and highly dimensional, Lasso Regression aggressively sets coefficients to zero if it cannot find a subset of variables that collectively explain the dependent variable. In this case, since the clinical scores are assessed empirically and lack strong correlations with the radiomics, this shrinkage method fails to identify any radiomic features as significant predictors for five out of the nine considered metrics.

TABLE 4:

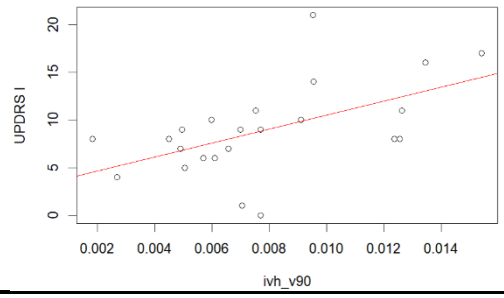
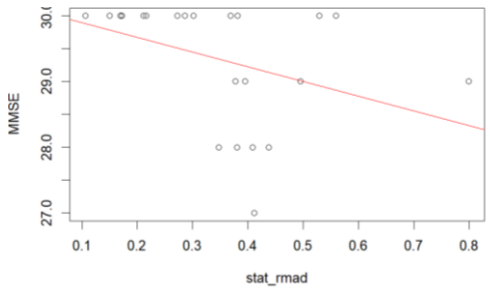
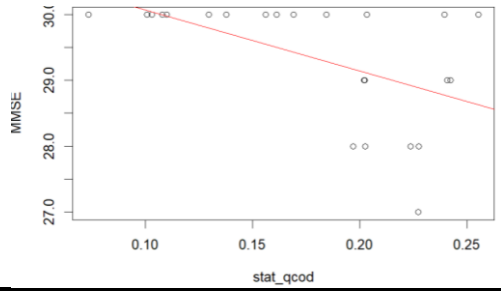
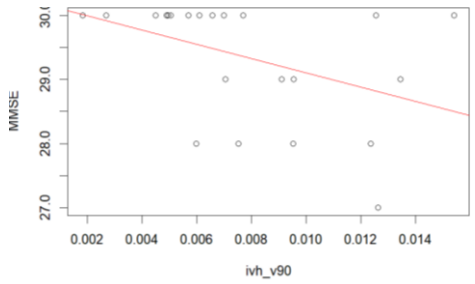
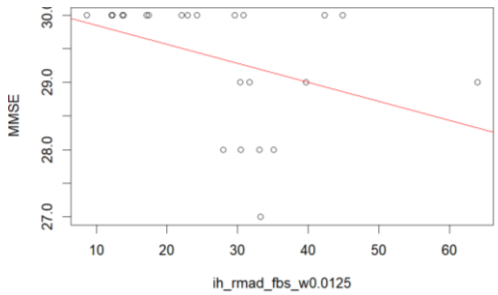
METRIC	RADIOMIC FEATURE	PLOT
UPDRS I	**ivh_v90:** Correlation: 0.4139 p-value: 0.0067	
MMSE	**stat_rmad:** Correlation: -0.4209 p-value: 0.0127	
	stat_qcod: Correlation: -0.4312 p-value: 0.0107	
	ivh_v90: Correlation: -0.4004 p-value: 0.0178	
	ih_rmad_fbs_w0.0125: Correlation: -0.4209 p-value: 0.0127	

Table 4 shows the correlations among the evaluation metrics and the radiomics features which were above the 0.4 chosen threshold. Correlation and p-values are displayed, while in the right column the values of the distributions are plotted against each other, with the minimum least square line.

However, it is worth noting that Lasso Regression did return selected features for the variables "LEDD_TOTAL," "UPDRS_I," "UPDRS_III," and "MMSE." Considering these metrics, it is possible to notice the presence of UPDRS_I and MMSE, which are also the variables in output of the correlation analysis.

The presence of the selected features for these specific clinical variables suggests a potential association between the radiomic features and these variables, albeit with a moderate correlation. Lasso Regression, as a feature selection method, recognizes this association and identifies important radiomic features for these specific clinical variables.

Furthermore, to develop a linear model encompassing all severity metrics, the Bottom-Up approach was employed. This method involved constructing a linear regression model for each score individually, selecting the most pertinent features according to its criteria.

Moreover, the results obtained from the previously conducted correlation analysis between the radiomics and PD severity metrics enabled us to perform linear regression between UPDRS_I and MMSE using their most correlated radiomics.

It is important to notice that the influence of the demographical information in the analysis was absolutely negligible. In fact, it affected only the prediction of the MoCA score with the Bottom-Up strategy and did it in a rather curious manner: only the variate 'height' was used in the regression and no radiomics feature was identified, making the prediction completely useless (RMSE = 0.6750, R-squared=0.1478, adjusted R-squared=0.1072, AIC=103.36).

To directly evaluate the influence of the demographic information on the severity assessment outcomes, we repeated the analysis taking into account only the radiomics features. All the implemented regression models with the two features selection methods returned the same outcomes, except for the ones of the MoCA score obtained with the Bottom-Up strategy: now the variable 'morph_pca_flatness' was identified as a regressor. All the outcomes and statistical values obtained from the latter linear regression analysis are presented in the table below (table 5).

Overall, the results obtained from the linear regression analysis show that we achieved relatively modest outcomes. This is not entirely unexpected considering the inherent challenges associated with explaining a subjective clinical variable using quantitative independent variables, the radiomics. The attempt to capture and predict clinical severity using radiomic features, although promising, faces inherent limitations. Quantitative variables may not fully capture the intricacies and nuances of subjective clinical assessments.

Nevertheless, upon examining the results, we observed that some of the radiomic features selected through the mentioned methods demonstrate a slight association with PD severity scores, as indicated by an adjusted R-squared above 0.30. Notably, the linear regression models constructed using the "UPDRS_I" metric as the dependent variable yielded the most effective results, with an adjusted R-squared of 0.5311 for Lasso regression and 0.4385 for the Bottom-Up method. In contrast, the "ivh_v90" feature, chosen through correlation analysis, exhibited a lower adjusted R-squared of 0.2577.

It is not surprising that the linear model built using the Bottom-Up method to describe the severity of Parkinson's associated with the "UPDRS_TOTAL" score should also be considered. This is because the "UPDRS_TOTAL" score encompasses the "UPDRS_I", "UPDRS_II", "UPDRS_III" and "UPDRS_IV" metrics, with the "UPDRS_I" metric playing a significant role in determining the overall score, yielding an adjusted R-squared of 0.48.

Tab 5.a

	LASSO FEATURES			BOTTOM-UP FEATURES		
LEDD_TOTAL	FEATURES SELECTED: "cm_info_corr2_d1_3d_avg_fbs_w0.0125 ", " szm_lzlge_3d_fbs_w0.012"	RMSE:	24.2789	FEATURES SELECTED: "cm_info_corr1_d1_3d_avg_fbs_w0.0125"	RMSE:	28.4654
		R squared:	0.3211		R squared:	0.2144
		R squared adjusted:	0.2533		R squared adjusted:	0.1770
		AIC:	335.7427		AIC:	337.1022
UPDRS_I	FEATURES SELECTED: "stat_mean", "ivh_v75", "ivh_v90 ", "cm_inv_diff_mom_norm_d1_3d_avg_fbs_w0.0125"	RMSE:	0.1178	FEATURES SELECTED: "ivh_v50" , "ivh_v75" , "morph_integ_int"	RMSE:	0.1632
		R squared:	0.6163		R squared:	0.5151
		R squared adjusted:	0.5311		R squared adjusted:	0.4385
		AIC:	126.4130		AIC:	129.7988
UPDRS_II				FEATURES SELECTED: "loc_peak_loc", "stat_var", "morph_poa_min_axis"	RMSE:	2.005
					R squared:	0.4522
					R squared adjusted:	0.3657
					AIC:	139.1364
UPDRS_III	FEATURES SELECTED: "morph_diam", "morph_poa_maj_axis", "morph_vol_dens_aabb ", "szm_lze_3d_fbs_w0.0125", "szm_lzlge_3d_fbs_w0.0125", "dzm_ldlge_3d_fbs_w0.0125"	RMSE:	6.1863	FEATURES SELECTED: "morph_diam", "cm_joint_max_d1_3d_avg_fbs_w0.0125"	RMSE:	6.3509
		R squared:	0.5045		R squared:	0.4203
		R squared adjusted:	0.3187		R squared adjusted:	0.3623
		AIC:	183.7286		AIC:	0.3623
UPDRS_IV				FEATURES SELECTED: "ih_min_grad_g_fbs_w0.0125" "szm_zs_var_3d_fbs_w0.0125"	RMSE:	2.3395
					R squared:	0.3247
					R squared adjusted:	0.2572
					AIC:	109.6988
UPDRS_TOTAL				FEATURES SELECTED: "morph_diam" "cm_inv_diff_mom_norm_d1_3d_avg_fbs_w0.0125" "szm_lze_3d_fbs_w0.0125", "szm_lgze_3d_fbs_w0.0125"	RMSE:	11.9501
					R squared:	_0.5753
					R squared adjusted:	0.4809
					AIC:	198.3590
NMSQ				FEATURES SELECTED: "ivh_v90"	RMSE:	0.7798
					R squared:	0.1959
					R squared adjusted:	0.1576
					AIC:	126.2715
MMSE	FEATURES SELECTED: "stat_qcod", "ivh_v75"	RMSE:	0.0665	FEATURES SELECTED: "stat_qcod", "dzm_sdlge_3d_fbs_w0.0125"	RMSE:	0.3917
		R squared:	0.3464		R squared:	0.4427
		R squared adjusted:	0.2811		R squared adjusted:	0.3870
		AIC:	59.3548		AIC:	55.6890
MoCA				FEATURES SELECTED: "morph_poa_flatness"	RMSE:	0.2364
					R squared:	0.1289
					R squared adjusted:	0.0874
					AIC:	103.8722

The table presents the results of the linear regression of a subset of features for different Parkinson's severity indices.

Tab 5.b

	"ivh_v90"		"stat_rmad", "stat_qcod", "ivh_v90" , "ih_rmad_fbs_w0.0125"	
UPDRS_I	RMSE:	6.1071		
	R squared:	0.2914		
	R squared adjusted:	0.2577		
	AIC:	134.5254		
MMSE			RMSE:	29.6532
			R squared:	0.4045
			R squared adjusted:	0.2721
			AIC:	61.2180

The table presents the results of the linear regression of a subset of features for different Parkinson's severity indices.

Discussion

The aim of the report is to provide a justified opinion on the use of radiomics analysis for DAT SPECT imaging as biomarker of Parkinson's Disease.

The first part of our analysis aimed at finding radiomics features capable of distinguishing patients from controls. From our findings, in particular looking at the classifiers' outcomes, it emerged that the best performing group of selected radiomics is the one obtained by the Lasso feature selection. In fact, four out of the six implemented classifiers achieved perfect accuracy; the remaining two still provided accuracy higher than 90% with a sensitivity equal to 1, being able to perfectly recognize the ill patients.

Nevertheless, a consideration must be made looking at the remarkable results obtained also with the Bottom-Up features: all the classifiers show 93% or higher accuracy, but the number of features in this case is two instead of eleven. More specifically, the selected bottom-up features represent the mean value of the grey levels of the DAT SPECT image (stat_mean) and the skewness of the intensity histogram of the image (stat_skew), two measures that are easily derivable from the scans; on the contrary, in the Lasso selected features are present radiomics that are extracted from more specific and advanced analysis, such as Region of Interest identification (ROIs) and Principal Component Analysis (PCA). This is an important factor to consider especially when dealing with the translational aspect of our study: having fewer features to compute and analyzing them in an easier way can significantly speed up the analysis process and make it more feasible in a clinical environment, not only saving time but also enhancing the practical applicability of our study's results. For this reason, choosing the Bottom-Up features could be a strategical choice that allows almost perfect classification with simpler features extraction procedures.

From the classification methods' prospective, the outcomes of the SVM and KNN classifiers outperform the other ones in terms of balanced accuracy and sensitivity/specificity metrics with all groups of features, with slightly better results from the k-nearest-neighbors which is able to achieve higher performances with fewer features. This conclusion is also supported by the De-Long test done in the statistical analysis part, which confirmed that the two previously mentioned classifiers had statistically significant differences in the performances if compared to the other methods employed. Although these are promising results for our report, they should be received with caution, since the size of our dataset was really limited in terms of analyzed subjects, and therefore it did not allow to conclude much on the robustness and reliability of our results. However, from our previous considerations it is possible to conclude that the best combination of features selected and employed classification method would be the Bottom-Up features alongside the KNN classifier.

It is worth adding that, although correctly classifying individual patients from healthy controls is a good starting point, its practical value is limited [4], since those patients are already correctly classified via presumably simpler clinical examinations such as the metrics analyzed. However, these metrics for Parkinson's disease take into account only the symptoms presence and persistence, while the possibility of diagnosing the disease from objectively measured variables such as the radiomics could lead to a biologically based 'gold standard', thus reducing the variability due to the partial subjectivity of the analyzed evaluation methods.

However, the main clinical value of machine learning methods and the associated radiomics features, when meant to be used as biomarkers, would come from their ability to detect subtle imaging signatures before the disease is clinically detectable or, as tried in the second part of our analysis, to point out the severity of neurological disorders such as the Parkinson's disease in the single individuals.

The present report intends also to investigate whether there is any difference between controls and PD in radiomics features: based on the Wilcoxon analysis evidence, most of the features (159 out of 177) exhibit a statistically significant difference in the distributions of subjects with Parkinson's disease and healthy controls and thus demonstrate discriminatory power between them. Among these features, having in mind the aim of finding the most useful in terms of discriminative capability among PD and HC subjects, the radiomics that are more easily derivable from the PET scans are the optimal ones to be considered.

The second part of our analysis aimed to seek if there is an existing association between the radiomic features and Parkinson's disease clinical symptoms severity. From the findings in output to the correlation analysis and the linear regression methods, it is possible to conclude that the influence of the demographic information to the analysis was completely negligible. We found these results quite peculiar, since we expected a higher influence of some demographic info (such as the age of the subjects). In fact, it is well known that aging causes to the brain irreversible modifications in structure and functioning [5], therefore potentially influencing the radiomic values. Moreover, aging represents a risk factor for the outbreak of Parkinson's disease. However, the explanation of these findings could be that, also if age could be a factor influencing Parkinson's outbreak, maybe it does not influence the severity of the disease itself.

In our study, due to its categorical values, gender was not included. Nevertheless, based on our previous consideration on age, we could formulate the same hypothesis for the gender variate. In fact, also if literature provides evidence that more men than women are diagnosed with PD by a ratio of approximately 2:1 [6], the gender factor might as well directly influence only the Parkinson's outbreak and potentially its symptoms, and not the severity itself. However, to be relied on, this hypothesis must be tested and supported by statistical evidence, therefore we suggest considering it for future studies.

Our results highlight that "UPDRS_I" seems to be the most suitable clinical metric to achieve our goals. Furthermore, we can state that the linear models built using Lasso and Bottom-Up selecting features methods outperformed the model based on features identified through correlation analysis. More specifically, when examining the AIC and adjusted R-squared values, Lasso Linear Regression seemed to be the best choice for our purposes.

However, it must be kept in mind that the translation of radiomic features into meaningful clinical insights is a complex process, influenced by various factors, including the heterogeneity of the patient population, inter-observer variability in clinical assessments and the multifaceted nature of Parkinson's disease symptoms. Therefore, explaining a subjective clinical variable as the previously mentioned metric using quantitative independent variables (radiomics) can be challenging and may not yield highly reliable results. Nevertheless, achieving relatively good outcomes with the "UPDRS_I" score is encouraging. These findings highlight the complexities of capturing and predicting clinical severity using radiomic features and underscore the importance of further investigation and refinement of the models for better understanding and application on the neurological disorders.

Having presented all the above considerations, it is crucial to highlight the intrinsic limitation that cursed our analysis since the beginning: the number of subjects available for analysis is extremely limited if compared to the number of radiomics, making the dataset extremely wide. This case is quite common in these types of studies, but it is the root of many issues. First of all, the training and test set have small dimensionality: with 53 patients at the beginning of our analysis, the split ratio is 38/15, while with the 33 subjects with Parkinson's disease this number decreases to 23/10. This impacts the reliability and the robustness of the implemented classification methods' results: although we tried to tune the parameters via LOOCV in order to increase the stability, it is possible that with a different training set the outcomes might change a bit. Furthermore, with such a small sample to analyze, the influence of the random generation effects (such as the seed setting) is increased and added to the impact of the high heterogeneity that usually neurological disorders datasets already present. For these reasons, analysis on small datasets such as ours are a good starting point, but must be taken with caution until replicated in larger studies, since they might result in potentially spurious results that don't replicate well [4].

In terms of applicability of our findings as biomarkers for Parkinson's disease, it is worth reminding that a biomarker gains value when it is not only statistically performative but also clinically relevant. More into detail, a biomarker should meet an authentic demand in the clinical population, allowing for easier diagnosis and treatment decisions than the standards already present (if any), and should be easily accessible, practically feasible and cost effective [7]. In our case, as we already mentioned, the clinical value of a biomarker able to perfectly classify already identified Parkinson's disease patients is limited, while its worth increases when dealing with the possibility of assessing Parkinson's severity or monitoring treatment's response. In spite of this, our conclusion would be that our findings at the moment do not support the hypothesis that such identified radiomic variates could be used as a good performing biomarker in this field of applicability. This conclusion is motivated by many reasons:

- In our Parkinson's severity analysis, the best outcome was able to explain only 53,11% of the total regression, meaning that almost half of it is influenced by unexplainable factors such as the high heterogeneity among the subjects.
- Even if this is a modest but potentially promising result, it must be taken into consideration that the DAT SPECT features derive from an imaging method that can cause many side effects (and thus cannot be applied systematically on a subject), it is expensive and not widely available in all hospitals and healthcare facilities. Therefore, the usual severity metrics, based on questions and visual examination done by the physician to the subject, result being, while approximative, implementable in an easier and quicker way. Despite of this, it is important to remember that in many cases the visual examination of SPECT images is traditionally applied to assess Parkinson's severity, in addition to the evaluation metrics analyzed in this study: in that case the use of the radiomics would be a non-indifferent improvement. Otherwise, it would be ideal to find a biological biomarker that does not rely on such imaging method, for the afore-mentioned motivations.

However, the lack of an objective and biologically based golden standard to evaluate neurological disorders such as the Parkinson's disease is a major issue for the clinical world, due to all the throwbacks deriving from the partial subjectivity of the actual evaluation metrics. Therefore, the research should move forward in the direction of finding a potential biomarker useful to address this issue. Our results, despite being modest, could be a useful starting point for future analysis, in particular for studies taking into account a broader population sample. In fact, due to the low dimensionality of our dataset, we may not have perceived relevant population dynamics which could have provided relevant insights on the complex mechanisms behind the Parkinson's disease. Moreover, our hope is that, thanks to future research on the potential of radiomics as clinically valuable biomarkers, the biomedical world would not only be able to assess disease's severity but also be capable of identifying early signs of the illness in order to perform a timely diagnosis.

Bibliography

[1] Rahmim, A., Salimpour, Y., Jain, S., Blinder, S. A., Klyuzhin, I. S., Smith, G. S., Mari, Z., & Sossi, V. (2016). Application of texture analysis to DAT SPECT imaging: Relationship to clinical assessments. *NeuroImage. Clinical*, *12*, e1–e9.

[2] Shiiba, T., Takano, K., Takaki, A., & Suwazono, S. (2022). Dopamine transporter single-photon emission computed tomography-derived radiomics signature for detecting Parkinson's disease. *EJNMMI research*, *12* (1), 39.

[3] https://molecularneuroimaging.wordpress.com/neuroimaging-database/

[4] Davatzikos C. (2019). Machine learning in neuroimaging: Progress and challenges. *NeuroImage*, *197*, 652–656.

[5] Mattson, M. P., & Arumugam, T. V. (2018). Hallmarks of Brain Aging: Adaptive and Pathological Modification by Metabolic States. *Cell metabolism*, *27*(6), 1176–1199.

[6] Miller, I. N., & Cronin-Golomb, A. (2010). Gender differences in Parkinson's disease: clinical characteristics and cognition. *Movement disorders: official journal of the Movement Disorder Society*, *25*(16), 2695–2703.

[7] Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular psychiatry*, *17*(12), 1174–1179.