A Data-driven Exploration and Prediction of Deep Brain Stimulation Effects on Gait in Parkinson's Disease

Gianluca Amprimo, *Member, IEEE*, Zhongke Mei, Claudia Ferraris, Gabriella Olmo, *Senior Member, IEEE*, Deepak K. Ravi

***Abstract*— Deep brain stimulation of the subthalamic nucleus (STN-DBS) is an established treatment for motor impairment due to Parkinson's disease (PD) progression. While treated subjects mostly experience significant amelioration of symptoms, some still report adverse effects. In particular, changes in gait patterns due to the electrical stimulation have shown mixed results across studies, with overall gait velocity improvement described as the core positive outcome. This retrospective study investigates changes in the gait parameters of 50 PD patients before and 6 months after STN-DBS, by exploiting a purely data-driven approach. First, unsupervised learning identifies clusters of subjects with similar variations in the gait parameters after STN-DBS. This analysis highlights two dominant clusters (Silhouette score: 0.45, Dunn index: 0.18), with one of them associated to a worsening in walking. Then, supervised machine learning models (i.e., Support Vector Machine and Ensemble Boosting models) are trained using pre-surgery gait parameters, clinical scores, and demographic information to predict the two gait change clusters. In a Leave-One-Subject-Out validation, the best model achieves balanced accuracy 80.05 $\pm$ 3.52 %, denoting moderate predictability of both clusters. Moreover, feature importance analysis reveals the variability in the step width and in the step length asymmetry during the preoperative gait test as promising biomarkers to predict gait response to STN-DBS.**

***Index Terms*— Deep Brain Stimulation, Parkinson's Disease, Gait, Machine Learning, Unsupervised Learning, Step Width, Gait Asymmetry**

I. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder which affects millions globally with progressive cognitive and motor impairment. The motor symptoms of PD are varied, including tremors, rigidity, bradykinesia, and postural instability, all of which profoundly impact daily functioning [1], [2]. In the context of PD axial symptoms, gait disturbances represent a significant challenge in daily life management, affecting mobility and independence. Gait in PD is commonly described as a slow, short-stepped, shuffling gait [3], [4], often associated to asymmetric lower limb movements [5], [6] and reduced bilateral coordination [7], [8]. While medication, especially Levodopa [9], remains the primary treatment in the early stages, its effectiveness may diminish over time, leading to fluctuations in symptom control and the onset of medication-related complications, such as dyskinesia [10], [11].

Deep brain stimulation of the subthalamic nucleus (STN-DBS) has emerged as a crucial intervention in managing PD symptoms, particularly those resistant to pharmacological treatment, such as tremor [12], [13]. The procedure involves the surgical implantation of electrodes into specific brain regions, with the subthalamic nucleus being a key target due to its involvement in motor control regulation [14]. By delivering high-frequency electrical impulses, the implanted electrode disrupts aberrant neural activity, restoring balance within the motor circuitry [15]. STN-DBS offers targeted and continuous stimulation, thereby alleviating symptoms, reducing fluctuations, and improving quality of life [16]–[18].

Although STN-DBS is widely accepted as an effective treatment, it is an invasive procedure and its success may be compromised by unpredictable efficacy limitations [19], [20]. These shortcomings include symptoms resistant to stimulation, such as postural instability, gait disorders, freezing unresponsive to medication, speech disturbances, psychiatric and cognitive dysfunctions, which may persist or worsen after surgery [20]–[23]. These aspects underscore the importance of a thorough patient selection, which is still mostly based on the criteria included in the CAPSIT-PD standard [24]. These criteria prioritize individuals with PD who are below the age range of 70–75, exhibit symptom reduction when on Levodopa (i.e., at least 30% improvement in Section III of the Movement Disorder Society's revision of the Unified Parkinsons's Disease Rating Scale (MDS-UPDRS)), possess minimal comorbidities or psychiatric conditions, and show no significant abnormalities on preoperative MRI brain scans. Moreover, eligible candidates typically exhibit motor fluctuations or dyskinesia,

This study was financially supported by grants from The LOOP Zürich and the Vontobel Foundation.

Gianluca Amprimo is with the Department of Control and Computer engineering of Politecnico di Torino, Torino, 10129, Italy (e-mail: gianluca.amprimo@polito.it) and CNR-IEIIT, Torino, 10129, Italy (e-mail:gianluca.amprimo@cnr.it)

Zhongke Mei is with the Laboratory for Movement Biomechanics, ETH Zürich, 8006, Zurich, Switzerland (e-mail: zhongke.mei@hest.ethz.ch)

Claudia Ferraris is with CNR-IEIIT, Torino, 10129, Italy (e-mail: claudia.ferraris@cnr.it)

Gabriella Olmo is with the Department of Control and Computer engineering of Politecnico di Torino, Torino, 10129, Italy (e-mail: gabriella.olmo@polito.it)

Deepak K. Ravi is with the Laboratory for Movement Biomechanics, ETH Zürich, 8006, Zurich, Switzerland (e-mail: deepak.ravi@hest.ethz.ch)

which persist despite pharmacological therapy [24], [25]. These selection criteria are quite restrictive: in a study by Morgante et al. [26] with 641 patients, the selection according to the CAPSIT-PD standard would have strictly included only 1.6% of the subjects, 4.5% when considering more relaxed constraints. Proposals for enlarging selection criteria include the addition of genetic screening, objective axial symptoms assessment through sensors, biomarkers related to nonmotor symptoms as well as measures of patients' expectation about the effect of the treatment [21], [27], [28].

While STN-DBS is generally acknowledged to improve gait velocity and thus confidence during walking, the overall response varies among individuals, with an estimated 25% of subjects experiencing exacerbation of gait abnormalities after the surgery [21], [29], [30]. These effects may translate, for instance, in an increased risk of falls, which are a significant cause of disability, lost independence and reduced quality of life in people with PD [31], [32]. This scenario highlights the need for precise characterization of gait alterations before STN-DBS and the identification of predictive markers to optimize patient outcomes in this motor domain.

This retrospective study investigates the changes in the gait of PD patients related to STN-DBS. In particular, the main novelty of the proposed approach relies in the use of machine learning (ML) methods for the identification of predictive biomarkers of STN-DBS effects on gait, which may be quantitatively measured before surgery. The first step to achieve this aim is an unsupervised exploration of the changes observed in the gait parameters collected through a motion capture system, on a cohort of 50 PD patients. Each participant was tested before and 6 months after STN-DBS surgery. The aim is to identify clusters of subjects with similar changes in gait parameters which may reflect different responses to the treatment. Then, the predictability of such clusters from only pre-surgical information (i.e., gait parameters as well as clinical scores) is explored by employing supervised ML techniques, to identify possible response biomarkers. The trained models and the identified biomarkers may be integrated into the clinical selection of candidates for STN-DBS thus providing a more comprehensive screening tool for personalized healthcare treatment.

II. Background

A. Effects of STN-DBS on gait

A plethora of works have investigated the effects of STN-DBS on walking patterns before and after surgery, by exploiting motion capture systems [29], [33]–[39]. However, these studies often present mixed and contradictory results. Indeed, a substantial limitation is the complexity of recruiting participants for performing instrumented and long-term follow-up studies. Most research efforts investigate the changes in gait patterns using an instrumented gait analysis system before surgery and 3, 6 or 12 (at most) months after, with a sample size ranging from 5 to around 30 patients [29], [34]–[36], [40]. The dataset employed in this study, first described in [39], is one of the largest ones, with currently 50 recruited subjects who completed the 6-month protocol. While larger and longer-term retrospective studies exist [20], [41], [42], they only report changes in clinical scales, such as those in MDS-UPDRS [43]. However, gait is undercharacterised in such a scale, with only a qualitative subscore ranging from 0 (no impairment) to 4 (total impairment).

The most consistent outcome across quantitative studies is an improvement in walking speed, usually associated with an increase in stride/step length [30], [44]–[46], but poor responsiveness in terms of cadence. This change is regarded overall as a positive response to STN-DBS and appears to be even more evident when electrical stimulation is coupled with Levodopa [45], [47], [48]. However, some studies [33], [35], [49] have found an opposite trend, suggesting that these changes may primarily manifest during short follow-up periods (e.g., three months), then may diminish over longer observation windows. The main explanation for such behaviour could be the natural progression of the disease [18], [33], [41].

Additional reported positive effects include a reduction in double-stance duration, spatial foot position asymmetry, stride-to-stride variability, and inter-limb coordination, thus a more regular gait cycle [34], [38], [48], [50]–[52]. Regarding muscular coordination, Fasano et al. [51] found out that a reduction of stimulation voltage on the side opposite the leg with the longer step length reduced freezing of gait (FOG) by normalizing gait symmetry and coordination, as measured by Phase Coordination Index (PCI). Moreover, from the perspective of muscle synergies, Ghislieri et al. [34] in a study on 20 PD patients identified a reduced number of muscle synergies with respect to healthy controls both in pre and post surgery conditions. However, *muscular robustness* increased overall for PD subjects both 6 and 12 months after surgery.

As for to results on gait variability, they are more complex to interpret. While high values may be associated with ageing and neurodegenerative processes [53], [54], even healthy subjects exhibit some variability [55], [56]. Therefore, an excessive reduction may also be prognostic of pathological walking [55], [57]. Regarding asymmetry and coordination, some studies reported a persistent worsening of these gait features, especially in the lower limbs [29], [35], [37] which could be responsible for the recurrent FOG events and falls reported after surgery [51], [58].

From the described scenario, it emerges a lack of a comprehensive analysis of gait in STN-DBS, jointly considering several perspectives (e.g., asymmetry, coordination, variability). This lack may be the reason, together with the reduced sample sizes, behind the inconsistencies in the outcomes found across different studies.

B. Predictability of STN-DBS outcomes

Previous works have addressed the predictability of STN-DBS outcomes from preoperative information. Most of them do not employ quantitative motor parameters, but try to perform this task by leveraging only demographic information and clinical scores [22], [59]–[64]. Moreover, a lack of consensus exists in defining *good* and *bad* responses to the treatment. For instance, Habets et al. [59] proposed a linear regression model called DBS-PREDICT to predict

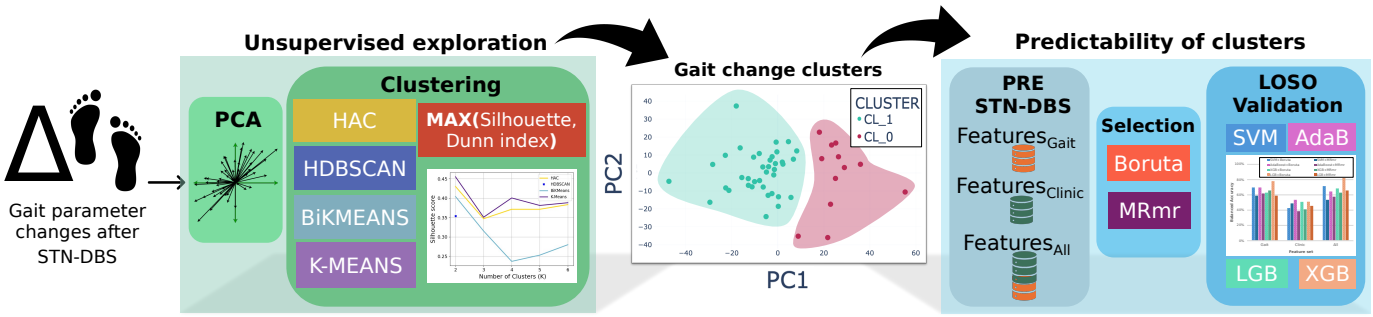

Fig. 1. Summary of the complete experiment: first, the identification of clusters of patients sharing similar gait changes between PRE and POST surgery. Second, the prediction of the identified clusters from preoperative-only data (gait features, clinical scores, and demographic data), to identify early biomarkers of the effects of STN-DBS on gait using feature selection methods and shallow learning models. Acronyms: PCA: Principal Component Analysis, HAC: Hierarchical Agglomerative Clustering, BikMEANS: Bisecting K-Means, MRmr: Maximum Relevance minimum redundancy, LOSO: Leave-One-Subject-Out.

probability of being *strong* or *weak* responders, i.e., subjects with a minimal clinical important difference (MCID) in either MDS-UPDRS II, III, or IV scores, one year after surgery. The model employs features such as patient's age at the time of surgery, disease duration, and clinical scores in Levodopa ON and OFF states. In a multi-centric study involving data from more than 300 patients (*weak* responders: 26%), the model achieved a 77% accuracy, confirming the preliminary results from their pilot study [23]. Also Frizon et al. [60], trained a logistic regression model to distinguish *good* and *bad* responders, but in terms of MCID in the quality of life (QoL), as measured by the Parkinson's Disease Questionnaire-39 [65]. Their model, considering features similar to Habets et al., achieved 81%, accuracy on a cohort of 77 patients (*bad* responders: 47%). Krause et al. [22] classified *good* responders as those with a decrease in the total score for MDS-UPDRS III, and *bad* responders those having no change or an increase. Then, preoperative demographic and clinical data were combined with supervised and semi-supervised ML models to predict these two groups. A maximum accuracy of 81.7% was achieved when classifying records from 105 patients (*bad* responders: 40%).

Studies performing predictions from quantitative motor parameters are under explored [35], [66]. Cebi et al. [35] preliminarly investigated the predictability of FOG outcomes 6 months after STN-DBS on 18 patients using multiple regression. The model included both clinical and demographic information, as well as parameters such as stride length, gait cycle time, walking speed, and swing time asymmetry, evaluated during a 7-meters timed walking test, repeated both in Levodopa ON and OFF states. Data were collected using three inertial measurement units, placed at both ankles and in lumbar position. Preoperative Levodopa sensitivity of FOG and ON/OFF variations of stride length and range of motion showed high correlation with a favourable outcome, also reflected by the regression model (R^2=0.952, p-value < 0.001). In [66], Shin et al. used Cox Proportional hazards analysis to predict relevant long-term clinical milestones (i.e., frequent falling, impaired walking, and loss of autonomy). They considered common spatiotemporal gait parameters retrospectively extracted from 96 video recordings of 63 PD patients during a 10 meters walking test. Variability in step length, step time, and stride time in ON condition were found predictive of all the three milestones.

The main limitation of the described works are related to their dependence on often arbitrarily and non univocal thresholds to establish *good* or *bad* response to STN-DBS, by leveraging clinical information that may be biased or not suitable for representing subtle gait changes. Moreover, most studies lack quantitative and objective measurements; this limits their generalizability, and introduces a subjective bias due to the use of qualitative clinical scales both as the ground truth and as predictors. This study, instead, expands the results in the direction of predicting changes after STN-DBS by employing quantitative gait data and provides a data-driven prediction, which does not depend on clinical scores and predefined thresholds to identify *good* and *bad* responders. Such definitions are therefore generated from the data themselves, through an initial unsupervised exploration of the changes observed between the preoperative and postoperative gait trials. Then, the prediction of the found clusters of response employs both qualitative and quantitative predictors to obtain relevant biomarkers of gait changes.

III. Materials & Methods

Fig. 1 summarises the complete experiment. The work articulates in two parts: first, the identification of clusters of patients sharing similar gait changes after surgery. Then, the prediction of the identified clusters from preoperative-only data, to identify early biomarkers of the effects of STN-DBS on gait.

A. STN-DBS dataset

For this study, the dataset previously collected and presented by Mei et al. in [39] is employed. The dataset contains a sample of 50 individuals with PD, who were recruited and assessed for gait kinematics, under ON medication, before STN-DBS electrode implantation (PRE), and approximately 6 months after (POST), under ON medication and ON stimulation condition. Inclusion criteria for patients were a diagnosis of idiopathic PD, an age in the 18-90 years old range, being designated for STN-DBS implantation at University Hospital

Zurich (USZ) and the ability of walking independently and continuously for 10 minutes. Data were collected with the approval of Kantonale Ethikkommission Zurich, Protocol Number: 2015-00141, and patients were recruited under informed consent. As described in previous studies using this dataset [29], [39], all participants walked barefoot for a continuous period of 10 minutes at a self-selected speed and without any assistance. Each participant was instructed to navigate an "8"-shaped path around two signs positioned 10 meters apart, to capture consecutive gait cycles during overground walking in the laboratory setting. The longer duration of the trials, with respect to similar studies in the literature, was chosen to obtain a more precise estimation of gait parameters and their variability, as suggested by previous studies such as [67]–[71]. While fatigue may occur in this scenario, none of the recruited subjects reported it at the end of the test, neither this phenomenon appeared in the collected data. Locomotion was recorded using a three-dimensional motion capture system (consisting of 10 cameras; 61 markers; sampling rate of 100 Hz; Vicon Nexus, version 2.3/2.8.2, Oxford Metrics, United Kingdom). Only straight path segments were recorded since the original protocol was designed to emulate treadmill evaluation for the assessment of continuous gait cycles. The turning phases were not recorded because they were outside the scope of the original study: in fact, given the peculiarity of turning in PD, this motor task was left for future ad hoc investigations. Additional clinical and demographics information (reported in the Supplementary Materials) were available from the electronic records of the clinical examinations of the patients, performed with the same timing of the gait trails (i.e., before and 6 months after surgery).

B. Gait parameters

The gait parameters included in this study were extracted using custom Matlab scripts (version R2022a, The MathWorks Inc., Natick). In further detail, mean, variability, and asymmetry of common spatiotemporal parameters were evaluated, as well as upper and lower limbs coordination. Spatiotemporal parameters include step length (StepL), step width (StepW), step time (StepT), stride length (StrideL), duration of the swing (SwingT), stance (StanceT) and double limb support (DLS) phases, cadence (Cad), and walking speed (WalkS). Asymmetry is evaluated as proposed in [72] for step length (ASYM_SL), step width (ASYM_SW), step time (ASYM_ST), swing phase (ASYM_SWG) and stance phase (ASYM_STC). Moreover, PCI from [7] and Continuous Relative Phase (CRP) from [7], [8], [73] measure coordination during walking; the former compares left and right (PCI_LR) and short and long gait cycles (PCI_SL); the latter compares coordination between upper limbs movements (CRP_ARMS), lower limbs movements (CRP_LEGS) and their combination (CRP_RA-LL, CRP_LA-RL, CRP_LA-LL, CRP_RA-RL).

Each parameter, including those for asymmetry and coordination, was evaluated in terms of mean value and coefficient of variation (CV), defined as (std/mean)*100. For spatiotemporal parameters, left and right gait cycles were averaged. Indeed, laterality is often a feature of PD, but the recognition of the most affected side may be non-univocal and wrongly bias the obtained results, especially in a dataset encompassing few subjects. To account for possible laterality patterns, asymmetry and coordination parameters were computed for each gait cycle and then averaged over the whole trial, after removing outliers, i.e., parameter values exceeding ±4 median absolute deviation (MAD) from the median.

C. Gait change clustering

Previous studies have applied clustering to gait data, for instance to identify walking patterns in patients with neurological conditions, PD included [74]–[78]. An unsupervised and data-driven investigation allows to identify information which may not be conveyed by subjective clinical scoring. Moreover, qualitative clinical scales such as MDS-UPDRS may not have sufficient granularity to detect small but relevant changes, as further discussed in Section IV.

In this work, the hypothesis is that clustering may highlight groups of subjects who exhibit a similar response to STN-DBS in terms of changes in their gait parameters. For each parameter P, its change ΔP is defined as $\Delta P = P_{POST} - P_{PRE}$, where PRE and POST refer to the values before and 6-months after the STN-DBS surgery. Such change is evaluated for mean values of spatiotemporal, asymmetry, and coordination parameters. Moreover, the change in the CV values of the spatiotemporal parameters is also included, to account for differences in the overall gait variability. The obtained dataset is thus characterised by 31 dimensions (gait parameter changes) for each patient. Principal component analysis (PCA) is applied for dimensionality reduction, by projecting this multidimensional points to the space identified by the main axis along which subjects are spread, easing the subsequent clustering. Indeed, clusters are computed considering only the principal components (PCs) with the highest contribution to the cumulative explained variance of the data, while discarding all the components which provide a negligible contribution. Furthermore, PCs allow to visualise such multidimensional data in a 2D or 3D representation space.

Four clustering methods are considered, namely Hierarchical Agglomerative Clustering (HAC), K-Means, Bisecting K-Means (BiKMeans), and Hierarchical DBSCAN (HDBSCAN) [79]. The maximisation of the Silhouette score (1) and of the Dunn index (2) drives the selection of the optimal clustering method and the optimal number of clusters K. Silhouette measures the quality of the clustering by calculating how similar a point is to its own cluster compared to the others and then averaging this result over the whole dataset. A score close to -1 stands for a meaningless grouping, whereas close to +1 means perfectly separated and compact groups of points.

$$\text{Silhouette} = \frac{1}{N}\sum_{i=1}^{N}\frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{1}$$

where $a(i)$ is the average distance from the ith data point to other points within the same cluster, $b(i)$ is the smallest average distance from the ith data point to points in a different cluster, and N is the total number of points.

Dunn index D is a clustering validation measure that evaluates the compactness and separation of clusters. It is defined as the ratio of the smallest inter-cluster distance to the largest intra-cluster distance. Bigger the value, better the clustering.

$$D = \frac{\min_{i \neq j}\{d(C_i, C_j)\}}{\max_{1 \leq k \leq K}\{d(C_k)\}} \tag{2}$$

where $d(C_i, C_j)$ represents the distance between clusters C_i and C_j, and $d(C_k)$ represents the diameter of cluster C_k.

The optimal clusters are investigated to highlight common characteristics in the patients who were grouped together. This procedure involves an analysis of the statistical distribution of the gait changes in each cluster and of the available demographic data and clinical scores. The latter did not contribute to the definition of PCs, but may highlight peculiarities in the patients with similar changes in gait (e.g., same gender, same age range or similar disease duration). Statistical testing, in the form of independent sample t-test, is employed for this analysis, using a 95% confidence level. Data normality is preliminarily assessed from Q-Q plots and by Shapiro-Wilk test, to employ the most suitable statistical method between Student's t test or Mann-Whitney U test for each parameter. All statistical analysis are conducted using the Jamovi tool [80] and Python 3.10 (*Scipy* library).

D. Gait change prediction

The second part of the analysis investigates the predictability of the found clusters by exploiting preoperative only data and supervised ML models. The code for this analysis was developed in Python 3.10. The hypothesis is that the gait parameters in the PRE STN-DBS surgery phase as well as demographic and clinical information, combined with feature selection and ML methods, may be able to predict how gait will change for a certain subject in the POST phase. Good results in terms of predictability could suggest the existence of movement biomarkers of gait response to STN-DBS, providing a new tool to support the decision-making process in STN-DBS candidate selection.

The proposed prediction pipeline consists of two main parts: first of all, the construction of candidate optimal feature sets for the prediction, by exploiting feature selection. Then, the training and validation of the prediction model using the obtained feature sets. From all the preoperative data available, three initial features sets are built:

- **Features$_{\text{Gait}}$** contains mean and CV values of all the parameters described in Subsection III-B from the PRE gait;
- **Features$_{\text{Clinic}}$** contains clinical data such as age at surgery, gender, age at disease onset, disease duration, Levodopa equivalent daily dose (LEDD), height, weight, PD type (i.e., posture and gait disorder (PIGD) or tremor dominant (TD)), total scores for Section I, II, IV of MDS-UPDRS, and total score of Section III of MDS-UPDRS, evaluated in Levodopa ON, Levodopa OFF and as variation ON/OFF (i.e., ON-OFF difference);
- **Features$_{\text{All}}$** is the union of the two previous sets.

Each feature set undergoes feature selection separately. Boruta and Maximum Relevance minimum redundancy (MRmr) algorithms are compared for this task. In particular, Boruta distinguishes relevant features by comparing their importance against randomly generated shadow features, facilitating selection of informative variables [81]. The Python implementation of Boruta in the *shap-hypetune* library was used, using Light Gradient Boosting (LGB) model as the base estimator and *shapley values* as the feature importance metric [82]. Shapley values are then used to rank the selected features. Conversely, MRmr maximizes relevance to the target variable while minimizing inter-correlation [83]. Boruta can automatically identify the optimal number of features, while MRmr's optimal feature set size is empirically set to five, to avoid overfitting since dealing with a small dataset.

After selection, six candidate feature sets are obtained. Each of them is used to train and validate four shallow learning classifiers for the prediction: Support Vector Machine (SVM), and three ensemble boosting models, namely Adaptive Boosting (AdaBoost), LGB, and eXtreme Gradient Boosting (XGB). The optimal hyperparameters for each model are found using Bayesian search [84], considering the search spaces reported in Supplementary Materials (Table 2). Prediction performance is obtained through a Leave-One-Subject-Out (LOSO) validation. In each iteration of this procedure, a patient is held out as the test set while the model is trained on the remaining patients. LOSO is particularly beneficial for small medical datasets as it ensures that the model is evaluated on entirely unseen patient data, thus providing a more realistic assessment of its generalization performance. Due to uneven size of the gait change clusters (further details in Section IV), balanced accuracy [85] is used to compare models' performance. For the optimal model, additional metrics such as precision, specificity, sensitivity, and F-1 score are also computed separately for each cluster and then averaged by weighting their contribution according to the size of the cluster [86]. Moreover, to improve the training phase, Synthetic Minority Over-sampling Technique (SMOTE) algorithm [87] is used to augment the minority class by creating synthetic training data samples.

Finally, in the perspective of identifying relevant biomarkers, the importance of each feature in the optimal set (i.e., the one resulting in the higher prediction accuracy) is investigated, by looking at the statistical correlation between the feature and the clustering label, and the relative importance (i.e., the absolute mean shapley value) assigned to the feature during the selection stage.

IV. Results and Discussion

A. Clustering results

PCA on the POST-PRE changes of gait parameters highlights two main directions of maximum variation (PC1, PC2) along which the data are spread, out of the 31 components obtained by the procedure. These first two PCs cumulatively explain more than 70% of the total variance, overshadowing the single contributions of the remaining 29 components,

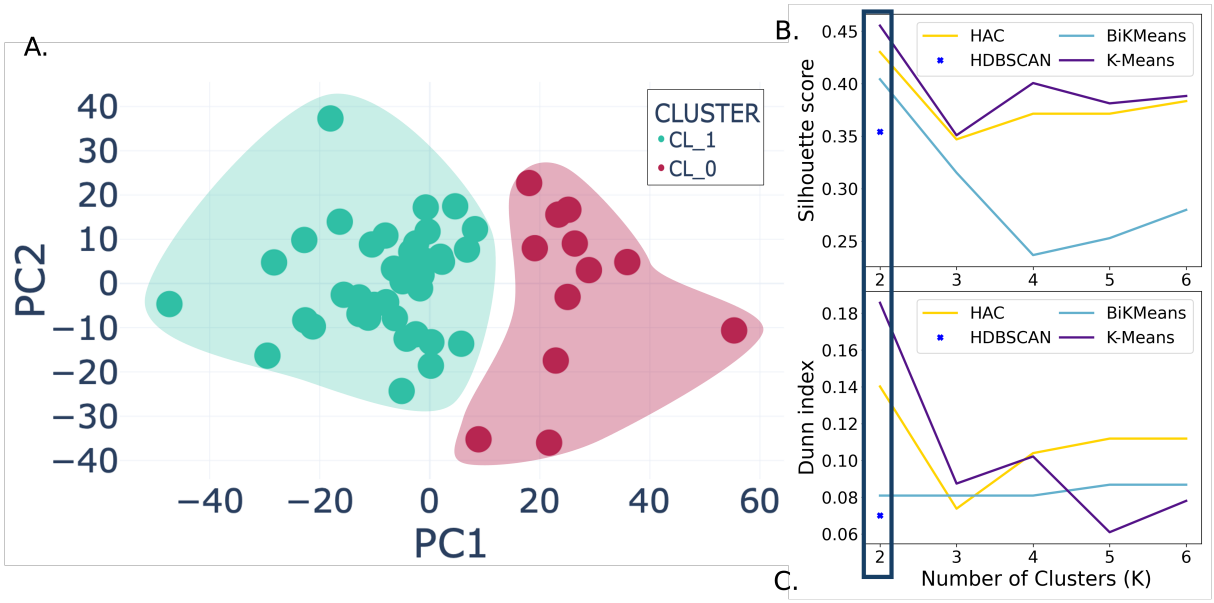

Fig. 2. Optimal clustering of patients based on the first two principal components (PC1 and PC2) from PCA, which summarize the variation in gait parameters following STN-DBS. Dunn index and Silhouette score are maximized for K=2 clusters, namely CL_0 (red) and CL_1 (green), identified using K-Means algorithm.

which either model noise or axes of minimal variability. Therefore, only PC1 and PC2 are retained for the clusterization. In addition, since PCs are linear combinations of the original gait parameter changes, from the coefficient associated to such changes their relative importance on the PCs can be found. This investigation highlights that $\Delta\text{StepW}_{\text{cv}}$, $\Delta\text{StepL}_{\text{mean}}$, $\Delta\text{StrideL}_{\text{mean}}$, $\Delta\text{WalkS}_{\text{mean}}$, and $\Delta\text{ASYM\_SW}_{\text{mean}}$ maximally contribute to the variance in the data. In other words, patients mostly differentiate one another after STN-DBS in terms of changes in gait velocity, spatial amplitude of lower limbs motion (step, stride), and variability and asymmetry in their step width.

Fig. 2 reports the results obtained by the clustering algorithms. The search for the optimal K was limited between 2 and 6 clusters due to the reduced number of points and their sparsity. All methods converge on $K = 2$ as the number of clusters which maximizes both Silhouette score and Dunn index. The optimal method results to be K-Means, with Silhouette score 0.46 and Dunn index 0.18. It is worth noting that HDBSCAN either identifies two clusters or classifies all points as noise or as a single cluster, when changing its hyperparameters. For these two conditions, however, neither Dunn index nor Silhouette score can be evaluated, thus a single point is reported in their graphs.

For convenience, the two clusters are named CL_0 and CL_1. Fig. 3 reports the box plots for changes of gait parameters which show a significant ($p < 0.05$) different statistical distribution between the two clusters. As expected, all the changes which were found significantly affecting PCs, namely $\Delta\text{StepW}_{\text{cv}}$ ($p < 0.001$), $\Delta\text{StepL}_{\text{mean}}$ ($p < 0.001$), $\Delta\text{StrideL}_{\text{mean}}$ ($p < 0.001$), $\Delta\text{WalkS}_{\text{mean}}$($p < 0.001$), and $\Delta\text{ASYM\_SW}_{\text{mean}}$ ($p < 0.001$) appear having a high significantly different distribution between CL_0 and CL_1. In addition, $\Delta\text{DLS}_{\text{mean}}$ ($p = 0.010$) also appears as statistically different. The most evident difference between the groups is the POST-PRE change of walking speed ($\Delta\text{WalkS}_{\text{mean}}$) and step/stride length ($\Delta\text{StepL}_{\text{mean}}$, $\Delta\text{StrideL}_{\text{mean}}$): CL_0 is characterised by a marked reduction of these gait parameters, whereas subjects in CL_1 appear to either maintain the same values or to increase them after STN-DBS, as expected from the combined use of electrical stimulation and Levodopa in the POST phase with respect to Levodopa only in the PRE one [47], [48]. Moreover, on average the double limb support phase (DLS_{mean}) gets longer for subjects in CL_0, whereas tends to reduce in CL_1.

Therefore, subjects in CL_1 have changes which are commonly associated to a *good* response to STN-DBS, as described by the literature review in Section II [30], [44]–[46]. On the contrary, patients in cluster CL_0 seem to worsen their gait patterns, therefore may model a *bad* response to the treatment. The postoperative reduction in step width variability ($\Delta\text{StepW}_{\text{cv}}$) and step width asymmetry ($\Delta\text{ASYM\_SW}_{\text{mean}}$) for CL_0, however, may appear contradictory with this interpretation of the two clusters. Nevertheless, as remarked in [39], the maintenance or slight increase in the value for step width variability (and its asymmetry) in cluster CL_1 may derive from the enhanced walking speed. Such improvement may require subjects to adapt more their base of support to preserve balance in a dynamic scenario, thus increasing step width variability and asymmetry. The decrease in CL_0, instead, may be associated to a more careful walking, because of a lack of perceived stability and fear of falling. This behaviour would also explain the longer double limb support duration (DLS_{mean}) for this group. In addition to these observations, the size unbalance between the two clusters (i.e., *bad* responders: 24%, *good* responders: 76%) appears coherent with the proportion of subjects who were reported not to experience benefit from STN-DBS in previous works [29], [59].

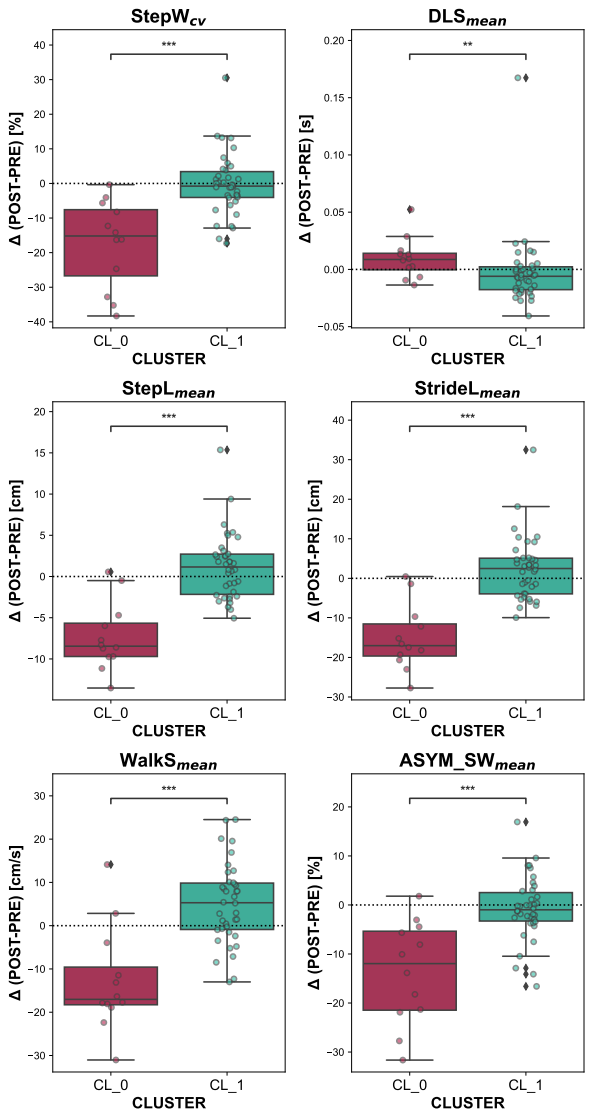

Fig. 3. Statistical characterization of the found clusters according to gait parameter changes (**Δ**) after STN-DBS surgery. Dotted line marks zero change between PRE-POST. Parameters were tested using either Student's t test or Mann-Whitney U test, according to their normality. *: p-value$< \mathbf{0.05}$, **: p-value$< \mathbf{0.01}$, ***: p-value$< \mathbf{0.001}$

The inspection of the two clusters in terms of age, age at disease onset, disease duration, height, weight, gender, LEDD variation, and MDS-UPDRS scores and their changes does not highlight any statistically significant difference ($p > 0.05$) between the two groups of patients (Supplementary Materials, Table 1). This result suggests that such factors may not have an influence on the gait changes discriminating the two groups. Furthermore, this suggests that this clinical information may not be predictive of the gait change cluster during the presurgery phase. This conclusion is further supported by the discussion in Subsection IV-B, which examines the predictability of the identified clusters from presurgery only data.

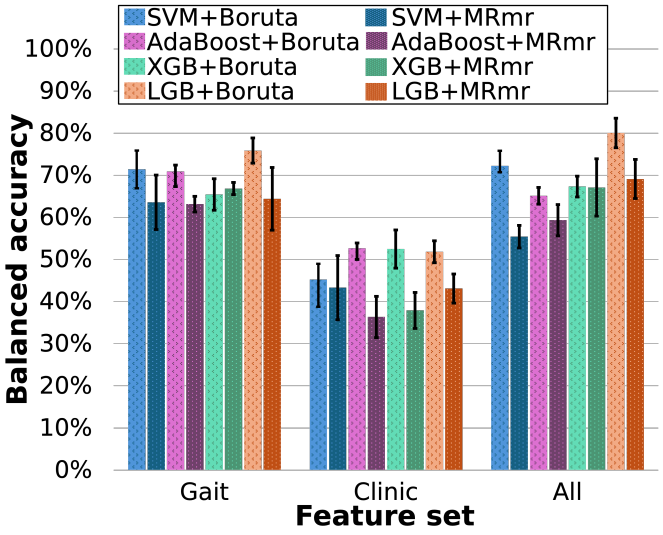

Fig. 4. Balanced accuracy achieved in the prediction of gait change clusters (CL_0, CL_1) from preoperative only data, organised in three feature sets ($\text{Features}_{\text{Gait}}$, $\text{Features}_{\text{Clinic}}$, $\text{Features}_{\text{All}}$). The best score is obtained by LGB model combined with Boruta as feature selection method, on $\text{Features}_{\text{All}}$. Bars represent mean values and error bars STD values computed on 5 Bayesian searches initialised using different random seeds.

B. Gait change prediction results

All the numerical results of this section are reported as $mean \pm std$. Such values were obtained by repeating the Bayesian search for each prediction pipeline (i.e., each combination of model, feature set and feature selection method) 5 times, using different random seeds for initialising the search. This procedure allows to provide a measure of performance variability and robustness to randomicity.

Fig. 4 reports the balanced accuracy obtained in LOSO validation by the investigated ML models and feature selection methods, for each investigated feature set ($\text{Features}_{\text{Gait}}$, $\text{Features}_{\text{Clinic}}$, $\text{Features}_{\text{All}}$). Regarding the latter, it can be observed how $\text{Features}_{\text{Clinic}}$ does not provide sufficient information to obtain a satisfactory prediction (accuracy between 40%-50%) independently from feature selection method and ML model. On the contrary, similar results are achieved for $\text{Features}_{\text{Gait}}$ and $\text{Features}_{\text{All}}$, suggesting that gait parameters from PRE phase likely provide most of the predicted power also in $\text{Features}_{\text{All}}$. The optimal estimation pipeline integrates LGB model and Boruta, on $\text{Features}_{\text{All}}$, with a balanced accuracy of 80.05 ± 3.52 %. This outcome is reasonable, considering that Boruta is a wrapper feature selection method specifically designed to exploit predictions from ensembles of decision trees.

Table I presents the precision, specificity, sensitivity, and F-1 score for this combination. As explained in Section IV, all the metrics are reported both per-class and as a weighted score (i.e., mean between the two prediction classes, with weights determined by the class proportion in the dataset). Examining the values, a gap in the prediction of the clusters appears. Despite the use of SMOTE and the maximisation of weighted F1-score in the Bayesian search for the best

TABLE I
LOSO VALIDATION PERFORMANCE OF THE BEST TRAINED MODEL (LGB MODEL WITH BORUTA, ON $\text{FEATURES}_{\text{ALL}}$)

Metric [%]	CL_0	CL_1	Weighted
Sensitivity	73.33 ± 3.73	86.84 ± 3.72	83.60 ± 3.58
Precision	64.21 ± 3.72	91.14 ± 1.45	84.67 ± 2.78
F-1 score	68.39 ± 5.60	88.92 ± 2.58	83.99 ± 2.91
Specificity	86.84 ± 3.72	73.33 ± 3.73	76.58 ± 3.58

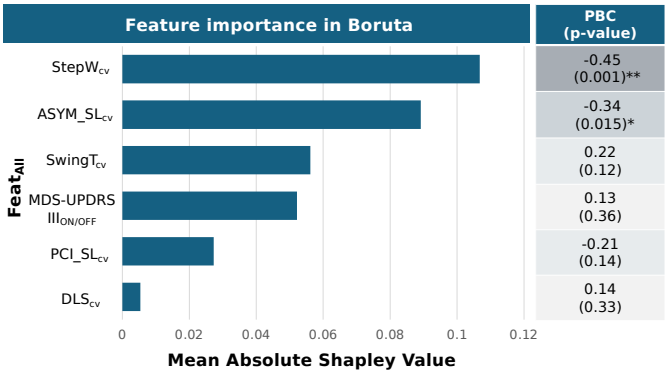

Fig. 5. Ranking of features in $\text{Features}_{\text{All}}$ during the feature selection procedure with Boruta algorithm, according to mean Shapley value. Moreover, PBC with cluster label is also reported, along its p-value, for each feature.

hyperparameters, the minority cluster results anyway hard to predict. Nevertheless, a good-to-excellent sensitivity is achieved for both clusters (CL_0: 73.33 ± 3.73 %, CL_1: 86.84 ± 3.72%), supporting that a high proportion of *good* and *bad* responders among the total is recognised (weighted sensitivity: 83.60 ± 3.58 %). Moreover, it must be noted that the high specificity obtained for CL_0 implies that the model has a low rate of false positives for this cluster. In the perspective of using the pipeline as a support system for selecting STN-DBS candidates, this means that unlikely *good* gait responders will be wrongly predicted as *bad*, lowering their chances of getting the intervention. The value of the weighted F-1 score (83.99 ± 2.91 %) suggests a good balance between sensitivity and precision. However, the precision for CL_0 (64.21 ± 3.72%) should be improved to reduce the number of subjects who are wrongly classified as CL_1. This low value may also depend on the very limited size of CL_0 (12 subjects).

Finally, while it must be remarked that a one-to-one comparison with other studies is not meaningful because of the different goals of the prediction (i.e., clinical motor scores vs gait change clusters), different validation methods (i.e., K-fold vs LOSO validation) and scoring metrics, the results obtained are in line and even slightly larger than those achieved in similar works on the prediction of STN-DBS effects, such as those by Habets et al. [23], [59] and Krause et al. [22].

C. Analysis of relevant predictors

The importance of the features which provided the best prediction results is further investigated. The optimal set results to be composed from six features picked up by the Boruta algorithm on $\text{Features}_{\text{All}}$, namely StepW_{cv}, $\text{ASYM\_SL}_{\text{cv}}$, $\text{SwingT}_{\text{cv}}$, $\text{MDS-UPDRS III}_{\text{ON/OFF}}$, $\text{PCI\_SL}_{\text{cv}}$ and DLS_{cv}.

First, it can be noted that almost all features refer to gait parameters in the PRE surgery phase, whose contribution in the prediction overshadows that of clinical and demographic information. Unsurprisingly, the only feature selected from these domains is $\text{MDS-UPDRS III}_{\text{ON/OFF}}$, in other words the change of motor condition, as evaluated by clinicians, between Levodopa ON and OFF states. This parameter is already considered as one of the fundamental selection criteria for STN-DBS. This result further support the importance of such criterion and conversely validate the relevance of the other selected features. Moreover, it must be noted that all the other features refer to CV values. This outcome seems coherent with the type of gait test that the subjects performed, which was specifically designed to address gait variability, due to the longer walking duration (around ten minutes), compared to other similar examinations (e.g., the 6-minute walking test). Furthermore, this result is coherent with previous findings by Shin et al. [66], who identified high predictability of postoperative clinical milestones from high preoperative gait variability in the Levodopa ON state.

Fig. 5 reports the ranking of the features according to Boruta algorithm, as derived from their mean absolute shapley values during the selection process. In addition, statistical correlation between such features and the gait change clusters is reported aside. Since correlation between continuous predictors and a binary outcome must be investigated, and all predictors exhibit a quasi-normal distribution, Point-Biserial Correlation (PBC) is employed [88], again with a 95% confidence. As shown, StepW_{cv} and $\text{ASYM\_SL}_{\text{cv}}$ appear to be the most relevant predictors, with also moderate correlation to the gait variation clusters, respectively -0.45 ($p = 0.001$), and -0.34 ($p = 0.015$).

Fig. 6 shows the distribution of the patients for these two preoperative parameters, coloured according to their gait change cluster. As it can be observed, on average, subjects in CL_0 are characterised by higher preoperative variability in step width and step length asymmetry, with a significant statistical difference for both features ($p < 0.01$). While high StepW_{cv} is known to be a marker of instability especially in the elderly [89], more complex is the interpretation of $\text{ASYM\_SL}_{\text{cv}}$, which is usually considered only in terms of mean value. Considering the protocol employed in data collection, variations in asymmetry in the straight walking segments may have been influenced by the turning events at the start and the end of the 8-shaped path. Therefore, $\text{ASYM\_SL}_{\text{cv}}$ may be influenced by residual effects of such turning events, suggesting that subjects who struggle to recover their normal walking pattern after turning may have higher values of $\text{ASYM\_SL}_{\text{cv}}$, thus higher probability to be *bad* responders for gait after STN-DBS.

D. Comparison with MDS-UPDRS changes

The exploration concludes with a comparison of the gait change clusters identified by the unsupervised learning approach, with the variations of MDS-UPDRS clinical scores.

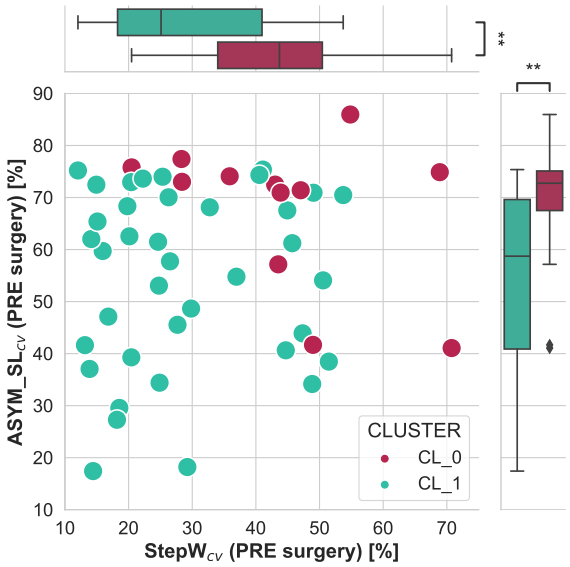

Fig. 6. Distribution of relevant predictive features (StepW$_{cv}$, ASYM_SL$_{cv}$) in the PRE surgery phase, colored according to POST surgery gait change clusters (CL_0 in red, CL_1 in green).*: p-value$<\mathbf{0.05}$, **: p-value$<\mathbf{0.01}$, ***: p-value$<\mathbf{0.001}$

Data were retrieved from electronic records of clinical examinations of each patient by a single neurologist. The comparison consider the preoperative scores in Levodopa ON state and those in STN-DBS ON and Levodopa ON, after surgery, thus the two conditions which were also considered in the gait trials. In particular, the analysis involves the changes, in the two conditions, of the MDS-UPDRS III score (i.e., the total motor score) and the MDS-UPDRS gait subscore. For the former, three groups of response are identified considering the definition from Horvath et al. [90]: *good* responders decreased their total score by at least 3.25 points; *bad* responders increased it by at least 4.6 points, and *non* responders fall in the range between these two values. For the gait subscore, a decrease by 1 point or more models a *good* response, conversely an increase by the same quantities a *bad* response and no change a *non* response. The gait subscore ranges between 0 (normal) and 4 (total impairment), and it must be pointed out that none of the patients in the cohort had gait score more than 2. This observation suggests that, on average, all subjects showed limited gait impairment when on pharmacological treatment only (PRE) or on electrical stimulation and reduced pharmacological therapy (POST), according to the clinical evaluator.

Fig. 7 provides a visualisation of the comparison of such groups with the identified clusters of gait change. The first evident result from the comparative analysis of the bars in the two plots is that an improvement in the total motor score ($n = 23$ patients) does not imply necessarily an improvement in the gait subscore, and viceversa. Most patients ($n = 30$) actually did not change their gait subscore in the two conditions. Only 11 patients had a worsening by exactly 1 point (i.e., either passing from score 0 to 1 or from 1 to 2), and only 9 an improvement by 1 point (i.e, from score 1 to 0), but these variations are not reflected by the total MDS-UPDRS III score. The total score involves the assessment of symptoms such as upper limb tremor, speech impairment, and facial expression whose improvement or worsening may not have any contribution at all on gait changes. Indeed, also the CL_0 and CL_1 are not reflected by response groups on MDS-UPDRS III, which supports the idea that, for an evaluation of the specific effects of STN-DBS on gait, just relying on predicting the variation of such score [22], [23] may not be relevant. A lack of coherence can be observed also between the response groups for the gait subscore and the gait change clusters. Several considerations should be taken into account. While the characterisation of the clusters appears coherent with previous description of positive and negative outcomes of STN-DBS on gait, such changes may not be *clinically relevant*. However, as discussed in the Background section, the definition of clinical relevance is often arbitrary and variable among different studies. This is also one of the reasons for selecting an initial unsupervised exploration to identify cluster of changes. More robust and long-term outcome measures such as quality of life questionnaires and scales specifically addressing gait, FOG and balance might have provided further insights into this discussion, but they were not part of the dataset employed in this retrospective investigation. Thus, proving evidence of clinical relevance should be addressed in future works. On the other hand, the well-know, limited granularity of the MDS-UPDRS subscores, with only five possible severity levels, may be the reason for this incoherence. This coarse staging can hardly reflect subtle improvements which may only be quantitatively measured by motion capture. Moreover, the longer test duration (10 minutes) of the gait trial in this study provides a more comprehensive insight into the walking impairment of the subject, which may not be observed by the clinician in the short duration of the standard neurological examination. Such longer duration could partially smooth out effects such as improved gait performance due to testing effect [91] and allow to better observe variability in gait parameters [67]. This overall outcome supports the importance of conducting an unsupervised exploration on quantitative parameters rather than relying only on variations of clinical scores, to really delve into the effects of STN-DBS on gait.

E. Limitations

Despite the promising results achieved, this work is not without limitations. As all the data-driven investigations, the statistical validity of the obtained results is limited by the observed sample size (50 patients), which however is much larger than those reported in previous equivalent studies (see Section II). This limited sample size may also explain why absolute changes were found to be more relevant for clusterization than percentage changes (with respect to PRE condition). While percentage changes may allow to individually evaluate improvement, they tended to make the data sparse for clustering, therefore were not considered. However, with larger samples available, they may provide further insights and allow to identify additional, finer-grained clusters.

Being a retrospective study, several limitations arise from the dataset itself. First of all, a *convenience* sample was

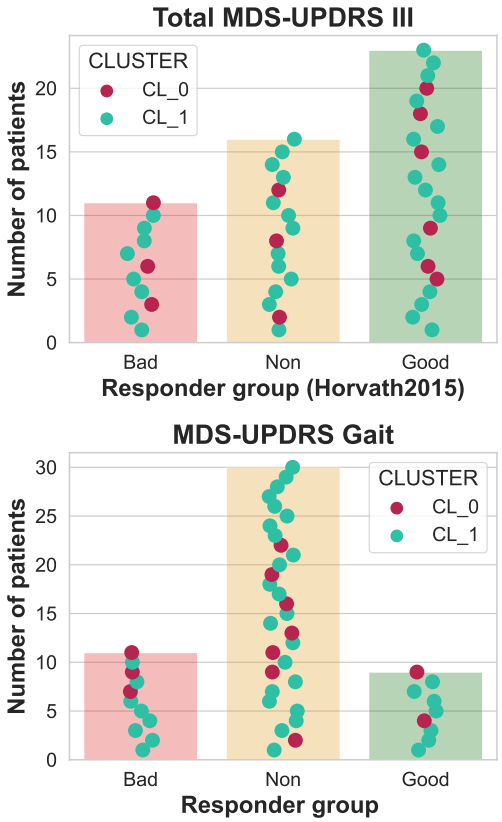

Fig. 7. Comparison of CL_0 and CL_1 with variations in Total MDS-UPDRS III score (top) and the gait subscore only (bottom). Clinical score variations are organised in responder groups, as defined in the literature [90].

selected for the original study, which may bias the results achieved in this work. In particular, since the capability of walking unassisted for 10 minutes was among inclusion criteria, this may have filtered out more severe subjects listed for STN-DBS as a late stage treatment. For severely impaired subjects, fatigue may be a relevant factor which could alter the observed gait parameters. Therefore, the obtained results may hardly generalize to this subgroup of patients. In addition, further information such as fall history and self-reported measures of gait quality, as well as cognitive data and scales measuring patient's independence in daily living could have provided a useful complement to the characterisation of the two found clusters. Indeed, lacking more robust and long-term scales, as mentioned in Section IV-D, does not allow to further investigate the clinical relevance of the observed gait changes. Moreover, patients included in the original study were not specifically recruited for treating gait impairment, which was only retrospectively considered. This also explains while on average most of the subjects were *good walkers* according to clinicians. Regarding this aspect, it must be noted that while many subjects with PD experience freezing of gait, the dataset of this study did not contain any subject with a clinical diagnosis of freezing neither freezing verified during the gait trails. Therefore, this lack may represent a further limitation to the generalizability of the results to a broader spectrum of patients.

Since Levodopa response is considered a main predictor of STN-DBS efficacy, motion capture data of gait trials in Levodopa OFF condition may have provided relevant biomarkers for the prediction of the two found clusters. Also, the small number of female patients (8/50) did not allow to completely rule out the existence of gender-related effects on the found gait changes after STN-DBS. Finally, the natural progression of the disease after the surgery was not considered as an influencing factor due to the complexity of modelling such a subjective aspect of PD, but for some patients may have represented a co-factor in the worsened gait patterns after STN-DBS.

F. Future developments

The current study is a first step towards the development of a tool that could support neurologists in selecting STN-DBS candidates. The ML model and the two found biomarkers (i.e., step width and step length asymmetry variability) may be used to leverage benefits and drawbacks of the treatment for the specific subject, prior to surgery. Indeed, the prediction of the gait change clusters could inform the clinicians about the likelihood of a worsening in walking characteristics, which may result in an increased risk of falls after surgery. On the contrary, the prediction of a maintenance or an improvement in walking capabilities could support the eligibility of the candidate.

Future studies should focus on enlarging the current observed sample size, which still represent a significant limitation. This would allow also to investigate the relevance of gait changes expressed as percentage change with respect to the PRE condition, in contrast to the absolute changes, chosen for this work due to the reduced sample size. Moreover, future studies should aim at generalising the obtained results to a more heterogeneous group of patients, including more severely impaired subjects. This would require to systematically model the effect of fatigue during the 10 minutes test, which would likely arise for this subgroup of patients. While collecting motion capture data may be a slow and expensive procedure, the use of wearable sensors (e.g., inertial measurement units) could be a compromise in accuracy to collect large amount of quantitative gait data. Moreover, wearable sensors could be used to perform assessment in real life conditions, for much longer periods than 10 minutes. Since variability appeared relevant in this investigation, a more continuous and pervasive monitoring could allow to obtain further insights into this domain. Especially the variability in the step length asymmetry should be further investigated in relation to walking patterns, to prove the hypothesis that this parameter may be associated to different immediate response to actions such as turning and its role in predicting effects of STN-DBS on gait.

Finally, in future studies, quantitative motion data should be better complemented by more robust and long-term clinical scales, which were missing in this dataset. Among clinical information, in-depth knowledge about dyskinesias events and time spent in OFF by the patients may provide enhanced characterization of clusters. Furthermore, it would be possible

to investigate whether reduced time in OFF due to STN-DBS continuous stimulation may influence gait performances and whether these two factors may be connected to post-surgery improvement.

V. Conclusions

This retrospective study investigated the effects of STN-DBS on gait in Parkinson's disease and their predictability, using machine learning methods. The core idea was to identify groups of patients who showed similar response in terms of changes in their gait parameters after surgery. For this exploration, motion capture data collected from 50 patients were employed, considering gait trials before and 6-months after surgery. The clustering of changes in gait parameters revealed two relevant groups. The statistical characterisation of the differences between the two clusters highlighted that the smaller cluster may represent a group of patients with a *bad* response on gait to STN-DBS, in contrast to the larger cluster which showed maintained or improved performance (i.e., a *good* response). The peculiarities of two clusters were coherent with findings from previous studies, and both clusters were predictable from preoperative only data, with a good level of accuracy ($80.05 \pm 3.52\%$). A final investigation of the most predictive preoperative features showed a connection between variability in step width and step length asymmetry and the gait change clusters.

Future works will focus on the consolidation of the obtained results (i.e., the found gait response groups and possible biomarkers), for the creation of a decision support system which could help clinicians in selecting optimal candidates for STN-DBS using quantitative gait parameters.

Acknowledgment

The authors extend their gratitude to all project members involved in the original data collection, including Dr. Navrag Singh, Prof. Dr. William Taylor, Dr. Christian Baumann, and their respective team members.

References

[1] A. A. Moustafa, S. Chakravarthy *et al.*, "Motor symptoms in parkinson's disease: A unified framework," *Neuroscience & Biobehavioral Reviews*, vol. 68, pp. 727–740, 2016.

[2] S. Rahman, H. J. Griffin *et al.*, "Quality of life in parkinson's disease: the relative importance of the symptoms," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 10, pp. 1428–1434, 2008.

[3] G. Ebersbach, C. Moreau *et al.*, "Clinical syndromes: Parkinsonian gait," *Movement Disorders*, vol. 28, no. 11, pp. 1552–1559, 2013.

[4] M. E. Morris, F. Huxham *et al.*, "The biomechanics and motor control of gait in parkinson disease," *Clinical biomechanics*, vol. 16, no. 6, pp. 459–470, 2001.

[5] F. Arippa, B. Leban *et al.*, "A study on lower limb asymmetries in parkinson's disease during gait assessed through kinematic-derived parameters," *Bioengineering*, vol. 9, no. 3, p. 120, 2022.

[6] B. W. Fling, C. Curtze, and F. B. Horak, "Gait asymmetry in people with parkinson's disease is linked to reduced integrity of callosal sensorimotor regions," *Frontiers in neurology*, vol. 9, p. 215, 2018.

[7] M. Plotnik, N. Giladi, and J. M. Hausdorff, "A new measure for quantifying the bilateral coordination of human gait: effects of aging and parkinson's disease," *Experimental brain research*, vol. 181, pp. 561–570, 2007.

[8] R. T. Roemmich, A. M. Field *et al.*, "Interlimb coordination is impaired during walking in persons with parkinson's disease," *Clinical Biomechanics*, vol. 28, no. 1, pp. 93–97, 2013.

[9] N. Tambasco, M. Romoli, and P. Calabresi, "Levodopa in parkinson's disease: current status and future developments," *Current neuropharmacology*, vol. 16, no. 8, pp. 1239–1252, 2018.

[10] T. Mueller and H. Russ, "Levodopa, motor fluctuations and dyskinesia in parkinson's disease," *Expert Opinion on Pharmacotherapy*, vol. 7, no. 13, pp. 1715–1730, 2006.

[11] B. Thanvi, N. Lo, and T. Robinson, "Levodopa-induced dyskinesia in parkinson's disease: clinical features, pathogenesis, prevention and treatment," *Postgraduate medical journal*, vol. 83, no. 980, pp. 384–388, 2007.

[12] M. Hariz and P. Blomstedt, "Deep brain stimulation for parkinson's disease," *Journal of internal medicine*, vol. 292, no. 5, pp. 764–778, 2022.

[13] J. K. Wong, J. H. Cauraugh *et al.*, "Stn vs. gpi deep brain stimulation for tremor suppression in parkinson disease: a systematic review and meta-analysis," *Parkinsonism & related disorders*, vol. 58, pp. 56–62, 2019.

[14] Y. Temel, A. Blokland *et al.*, "The functional role of the subthalamic nucleus in cognitive and limbic circuits," *Progress in neurobiology*, vol. 76, no. 6, pp. 393–413, 2005.

[15] W.-J. Neumann, L. A. Steiner, and L. Milosevic, "Neurophysiological mechanisms of deep brain stimulation across spatiotemporal resolutions," *Brain*, vol. 146, no. 11, pp. 4456–4468, 07 2023. [Online]. Available: https://doi.org/10.1093/brain/awad239

[16] W. M. Schuepbach, L. Tonder *et al.*, "Quality of life predicts outcome of deep brain stimulation in early parkinson disease," *Neurology*, vol. 92, no. 10, pp. e1109–e1120, 2019.

[17] C. Schlenstedt, A. Shalash *et al.*, "Effect of high-frequency subthalamic neurostimulation on gait and freezing of gait in parkinson's disease: a systematic review and meta-analysis," *European journal of neurology*, vol. 24, no. 1, pp. 18–26, 2017.

[18] H. Brozova, I. Barnaure *et al.*, "Short- and long-term effects of dbs on gait in parkinson's disease," *Frontiers in Neurology*, vol. 12, 2021. [Online]. Available: https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2021.688760

[19] S. Groiss, L. Wojtecki *et al.*, "Review: Deep brain stimulation in parkinson's disease," *Therapeutic Advances in Neurological Disorders*, vol. 2, no. 6, pp. 379–391, 2009. [Online]. Available: https://doi.org/10.1177/1756285609339382

[20] A. Williams, S. Gill *et al.*, "Deep brain stimulation plus best medical therapy versus best medical therapy alone for advanced parkinson's disease (pd surg trial): a randomised, open-label trial," *The Lancet Neurology*, vol. 9, no. 6, pp. 581–591, 2010.

[21] V. Geraedts, M. Kuijf *et al.*, "Selecting candidates for deep brain stimulation in parkinson's disease: the role of patients' expectations," *Parkinsonism & Related Disorders*, vol. 66, pp. 207–211, 2019.

[22] K. J. Krause, F. Phibbs *et al.*, "Predicting motor responsiveness to deep brain stimulation with machine learning," in *AMIA Annual Symposium Proceedings*, vol. 2021. American Medical Informatics Association, 2021, p. 651.

[23] J. G. Habets, M. L. Janssen *et al.*, "Machine learning prediction of motor response after deep brain stimulation in parkinson's disease—proof of principle in a retrospective cohort," *PeerJ*, vol. 8, p. e10317, 2020.

[24] G.-L. Defer, H. Widner *et al.*, "Core assessment program for surgical interventional therapies in parkinson's disease (capsit-pd)," *Movement disorders: official journal of the Movement Disorder Society*, vol. 14, no. 4, pp. 572–584, 1999.

[25] P. Limousin and T. Foltynie, "Long-term outcomes of deep brain stimulation in parkinson disease," *Nature Reviews Neurology*, vol. 15, no. 4, pp. 234–242, 2019.

[26] L. Morgante, F. Morgante *et al.*, "How many parkinsonian patients are suitable candidates for deep brain stimulation of subthalamic nucleus? results of a questionnaire," *Parkinsonism & related disorders*, vol. 13, no. 8, pp. 528–531, 2007.

[27] C. A. Artusi, L. Lopiano, and F. Morgante, "Deep brain stimulation selection criteria for parkinson's disease: time to go beyond capsit-pd," *Journal of clinical medicine*, vol. 9, no. 12, p. 3931, 2020.

[28] M. G. Rizzone, T. Martone *et al.*, "Genetic background and outcome of deep brain stimulation in parkinson's disease," *Parkinsonism & related disorders*, vol. 64, pp. 8–19, 2019.

[29] D. K. Ravi, C. R. Baumann *et al.*, "Does subthalamic deep brain stimulation impact asymmetry and dyscoordination of gait in parkinson's disease?" *Neurorehabilitation and Neural Repair*, vol. 35, no. 11, pp. 1020–1029, 2021.

[30] M. Pötter-Nerger and J. Volkmann, "Deep brain stimulation for gait and postural symptoms in parkinson's disease," *Movement Disorders*, vol. 28, no. 11, pp. 1609–1615, 2013.

[31] A. Fasano, C. G. Canning *et al.*, "Falls in parkinson's disease: a complex and evolving picture," *Movement disorders*, vol. 32, no. 11, pp. 1524–1536, 2017.

[32] J. Pressley, E. Louis *et al.*, "The impact of comorbid disease and injuries on resource use and expenditures in parkinsonism," *Neurology*, vol. 60, no. 1, pp. 87–93, 2003.

[33] R. Tripathi, J. L. McKay *et al.*, "Impact of deep brain stimulation on gait in parkinson disease: A kinematic study," *Gait Posture*, vol. 108, pp. 151–156, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0966636223015060

[34] M. Ghislieri, M. Lanotte *et al.*, "Muscle synergies in parkinson's disease before and after the deep brain stimulation of the bilateral subthalamic nucleus," *Scientific Reports*, vol. 13, no. 1, p. 6997, 2023.

[35] I. Cebi, M. Scholten *et al.*, "Clinical and kinematic correlates of favorable gait outcomes from subthalamic stimulation," *Frontiers in Neurology*, vol. 11, p. 212, 2020.

[36] D. W. Powell, S. E. Blackmore *et al.*, "Deep brain stimulation enhances movement complexity during gait in individuals with parkinson's disease," *Neuroscience letters*, vol. 728, p. 133588, 2020.

[37] N. Allert, J. Volkmann *et al.*, "Effects of bilateral pallidal or subthalamic stimulation on gait in advanced parkinson's disease," *Movement disorders: official journal of the Movement Disorder Society*, vol. 16, no. 6, pp. 1076–1085, 2001.

[38] P. Krystkowiak, J.-L. Blatt *et al.*, "Effects of Subthalamic Nucleus Stimulation and Levodopa Treatment on Gait Abnormalities in Parkinson Disease," *Archives of Neurology*, vol. 60, no. 1, pp. 80–84, 01 2003. [Online]. Available: https://doi.org/10.1001/archneur.60.1.80

[39] Z. Mei, A.-S. Hofer *et al.*, "Optimal stimulation sites of the subthalamic nucleus for the treatment of gait symptoms of parkinson's disease," *medRxiv*, 2023. [Online]. Available: https://www.medrxiv.org/content/early/2023/12/18/2023.12.15.23299998

[40] R. Roemmich, J. A. Roper *et al.*, "Gait worsening and the microlesion effect following deep brain stimulation for essential tremor," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 90, no. 8, pp. 913–919, 2019.

[41] P. Krack, A. Batir *et al.*, "Five-year follow-up of bilateral stimulation of the subthalamic nucleus in advanced parkinson's disease," *New England Journal of Medicine*, vol. 349, no. 20, pp. 1925–1934, 2003.

[42] G. Deuschl, C. Schade-Brittinger *et al.*, "A randomized trial of deep-brain stimulation for parkinson's disease," *New England Journal of Medicine*, vol. 355, no. 9, pp. 896–908, 2006.

[43] C. G. Goetz, B. C. Tilley *et al.*, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.

[44] J. A. Roper, N. Kang *et al.*, "Deep brain stimulation improves gait velocity in parkinson's disease: a systematic review and meta-analysis," *Journal of neurology*, vol. 263, pp. 1195–1203, 2016.

[45] A. Collomb-Clerc and M.-L. Welter, "Effects of deep brain stimulation on balance and gait in patients with parkinson's disease: A systematic neurophysiological review," *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 45, no. 4, pp. 371–388, 2015, special issue : Balance and Gait. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0987705315000404

[46] R. Pourahmad, K. Saleki *et al.*, "Deep brain stimulation (dbs) as a therapeutic approach in gait disorders: What does it bring to the table?" *IBRO Neuroscience Reports*, vol. 14, pp. 507–513, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667242123000465

[47] S. Lubik, W. Fogel *et al.*, "Gait analysis in patients with advanced parkinson disease: different or additive effects on gait induced by levodopa and chronic stn stimulation," *Journal of neural transmission*, vol. 113, pp. 163–173, 2006.

[48] J. M. Hausdorff, L. Gruendlinger *et al.*, "Deep brain stimulation effects on gait variability in parkinson's disease," *Movement disorders: official journal of the Movement Disorder Society*, vol. 24, no. 11, pp. 1688–1692, 2009.

[49] L. Rocchi, P. Carlson-Kuhta *et al.*, "Effects of deep brain stimulation in the subthalamic nucleus or globus pallidus internus on step initiation in parkinson disease," *Journal of neurosurgery*, vol. 117, no. 6, pp. 1141–1149, 2012.

[50] E. L. Johnsen, P. H. Mogensen *et al.*, "Improved asymmetry of gait in parkinson's disease with dbs: gait and postural instability in parkinson's disease treated with bilateral deep brain stimulation in the subthalamic nucleus," *Movement Disorders*, vol. 24, no. 4, pp. 588–595, 2009.

[51] A. Fasano, J. Herzog *et al.*, "Modulation of gait coordination by subthalamic stimulation improves freezing of gait," *Movement Disorders*, vol. 26, no. 5, pp. 844–851, 2011.

[52] M. Ferrarin, I. Carpinella *et al.*, "Unilateral and bilateral subthalamic nucleus stimulation in parkinson's disease: Effects on emg signals of lower limb muscles during walking," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 2, pp. 182–189, 2007.

[53] F. Pieruccini-Faria, M. Montero-Odasso, and J. M. Hausdorff, *Gait Variability and Fall Risk in Older Adults: The Role of Cognitive Function*. Cham: Springer International Publishing, 2020, pp. 107–138. [Online]. Available: https://doi.org/10.1007/978-3-030-24233-6_7

[54] F. Pieruccini-Faria, S. E. Black *et al.*, "Gait variability across neurodegenerative and cognitive disorders: Results from the canadian consortium of neurodegeneration in aging (ccna) and the gait and brain study," *Alzheimer's & Dementia*, vol. 17, no. 8, pp. 1317–1328, 2021.

[55] J. H. Richard E.A. Van Emmerik and W. J. McDermott, "Variability and coordinative function in human gait," *Quest*, vol. 57, no. 1, pp. 102–123, 2005.

[56] G. Y. Park, S. S. Yeo *et al.*, "Changes in gait parameters and gait variability in young adults during a cognitive task while slope and flat walking," *Healthcare*, vol. 8, no. 1, 2020. [Online]. Available: https://www.mdpi.com/2227-9032/8/1/30

[57] J. S. Brach, J. E. Berlin *et al.*, "Too much or too little step width variability is associated with a fall history in older persons who walk at or near normal gait speed," *Journal of neuroengineering and rehabilitation*, vol. 2, pp. 1–8, 2005.

[58] E. Lai, M. Bryant *et al.*, "Risk of falls in parkinson's disease after deep brain stimulation (p04.184)," *Neurology*, vol. 80, no. 7_supplement, pp. P04.184–P04.184, 2013. [Online]. Available: https://www.neurology.org/doi/abs/10.1212/WNL.80.7_supplement.P04.184

[59] J. G. Habets, C. Herff *et al.*, "Multicenter Validation of Individual Preoperative Motor Outcome Prediction for Deep Brain Stimulation in Parkinson's Disease," *Stereotactic and Functional Neurosurgery*, vol. 100, no. 2, pp. 121–129, 11 2021. [Online]. Available: https://doi.org/10.1159/000519960

[60] L. A. Frizon, O. Hogue *et al.*, "Quality of life improvement following deep brain stimulation for parkinson disease: development of a prognostic model," *Neurosurgery*, vol. 85, no. 3, pp. 343–349, 2019.

[61] H. S. Dafsari, L. Weiß *et al.*, "Short-term quality of life after subthalamic stimulation depends on non-motor symptoms in parkinson's disease," *Brain stimulation*, vol. 11, no. 4, pp. 867–874, 2018.

[62] C. Daniels, P. Krack *et al.*, "Is improvement in the quality of life after subthalamic nucleus stimulation in parkinson's disease predictable?" *Movement disorders*, vol. 26, no. 14, pp. 2516–2521, 2011.

[63] P. Voruz, J. Pierce *et al.*, "Motor symptom asymmetry predicts non-motor outcome and quality of life following stn dbs in parkinson's disease," *Scientific reports*, vol. 12, no. 1, p. 3007, 2022.

[64] O. Gavriliuc, S. Paschen *et al.*, "Prediction of the effect of deep brain stimulation on gait freezing of parkinson's disease," *Parkinsonism Related Disorders*, vol. 87, pp. 82–86, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1353802021001395

[65] C. Jenkinson, R. Fitzpatrick *et al.*, "The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score," *Age and Ageing*, vol. 26, no. 5, pp. 353–357, 09 1997. [Online]. Available: https://doi.org/10.1093/ageing/26.5.353

[66] J. H. Shin, R. Yu *et al.*, "High preoperative gait variability is a prognostic predictor of gait and balance in parkinson disease patients with deep brain stimulation," *Parkinsonism Related Disorders*, vol. 100, pp. 1–5, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1353802022001444

[67] N. König, N. B. Singh *et al.*, "Is gait variability reliable? an assessment of spatio-temporal parameters of gait variability during continuous overground walking," *Gait & posture*, vol. 39, no. 1, pp. 615–617, 2014.

[68] J. M. Hausdorff, M. E. Cudkowicz *et al.*, "Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in parkinson's disease and huntington's disease," *Movement Disorders*, vol. 13, no. 3, pp. 428–437, 1998. [Online]. Available: https://movementdisorders.onlinelibrary.wiley.com/doi/abs/10.1002/mds.870130310

[69] E. Sejdić, B. Findlay *et al.*, "The effects of listening to music or viewing television on human gait," *Computers in Biology and Medicine*, vol. 43, no. 10, pp. 1497–1501, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482513001935

[70] B. Galna, S. Lord, and L. Rochester, "Is gait variability reliable in older adults and parkinson's disease? towards an optimal testing protocol," *Gait Posture*, vol. 37, no. 4, pp. 580–585, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0966636212003712

[71] D. K. Ravi, M. Gwerder *et al.*, "Revealing the optimal thresholds for movement performance: A systematic review and meta-analysis to benchmark pathological walking behaviour," *Neuroscience Biobehavioral Reviews*, vol. 108, pp. 24–33, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0149763419303951

[72] W. Nanhoe-Mahabier, A. Snijders *et al.*, "Walking patterns in parkinson's disease with and without freezing of gait," *Neuroscience*, vol. 182, pp. 217–224, 2011.

[73] X. Huang, J. M. Mahoney *et al.*, "Both coordination and symmetry of arm swing are reduced in parkinson's disease," *Gait & posture*, vol. 35, no. 3, pp. 373–377, 2012.

[74] S. Mulroy, J. Gronley *et al.*, "Use of cluster analysis for gait pattern classification of patients in the early and late recovery phases following stroke," *Gait & posture*, vol. 18, no. 1, pp. 114–125, 2003.

[75] H. Kim, Y.-H. Kim *et al.*, "Pathological gait clustering in post-stroke patients using motion capture data," *Gait Posture*, vol. 94, pp. 210–216, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0966636222000790

[76] E. Dolatabadi, A. Mansfield *et al.*, "Mixture-model clustering of pathological gait patterns," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1297–1305, 2017.

[77] H. Zhao, J. Xie, and J. Cao, "Early detection of parkinson's disease by unsupervised learning from plantar bend data," 2022, Conference paper, cited by: 2. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85137783996&doi=10.1109%2fCIVEMSA53371.2022.9853703&partnerID=40&md5=2b4d0ce943ce2ab95b4b0a0772e6eb2b

[78] M. Serrao, G. Chini *et al.*, "Identification of specific gait patterns in patients with cerebellar ataxia, spastic paraplegia, and parkinson's disease: A non-hierarchical cluster analysis," *Human Movement Science*, vol. 57, pp. 267–279, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167945717300891

[79] C. X. Gao, D. Dwyer *et al.*, "An overview of clustering methods with guidelines for application in mental health research," *Psychiatry Research*, vol. 327, p. 115265, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165178123002159

[80] J. developers, "Jamovi - stats. open. now," http://www.jamovi.org, accessed: 2022-8-6.

[81] M. B. Kursa, A. Jankowski, and W. R. Rudnicki, "Boruta–a system for feature selection," *Fundamenta Informaticae*, vol. 101, no. 4, pp. 271–285, 2010.

[82] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg *et al.*, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[83] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*, 2003, pp. 523–528.

[84] J. Wu, X.-Y. Chen *et al.*, "Hyperparameter optimization for machine learning models based on bayesian optimizationb," *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1674862X19300047

[85] J. D. Kelleher, B. Mac Namee, and A. D'arcy, *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press, 2020.

[86] G. S. Handelman, H. K. Kok *et al.*, "Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods," *American Journal of Roentgenology*, vol. 212, no. 1, pp. 38–43, 2019.

[87] N. V. Chawla, K. W. Bowyer *et al.*, "Smote: synthetic minority oversampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[88] D. Kornbrot, "Point biserial correlation," *Wiley StatsRef: Statistics Reference Online*, 2014.

[89] A. Skiadopoulos, E. E. Moore *et al.*, "Step width variability as a discriminator of age-related gait changes," *Journal of NeuroEngineering and Rehabilitation*, vol. 17, pp. 1–13, 2020.

[90] K. Horváth, Z. Aschermann *et al.*, "Minimal clinically important difference on the motor examination part of mds-updrs," *Parkinsonism Related Disorders*, vol. 21, no. 12, pp. 1421–1426, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1353802015300079

[91] M. M. Ardestani and T. G. Hornby, "Effect of investigator observation on gait parameters in individuals with stroke," *Journal of Biomechanics*, vol. 100, p. 109602, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021929020300051