Gait & Posture 98 (2022) 49–55

Contents lists available at ScienceDirect

Gait & Posture

journal homepage: www.elsevier.com/locate/gaitpost

Machine learning models for Parkinson's disease detection and stage classification based on spatial-temporal gait parameters

Marta Isabel A.S.N Ferreira [a], Fabio Augusto Barbieri [a,*], Vinícius Christianini Moreno [b], Tiago Penedo [b], João Manuel R.S. Tavares [a,c]

[a] Faculdade de Engenharia, Universidade do Porto, Portugal
[b] São Paulo State University (Unesp), Department of Physical Education, Human Movement Research Laboratory (MOVI-LAB), Bauru, Brazil
[c] Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Portugal

ARTICLE INFO

Keywords:
Parkinson's disease
Classification
Algorithm
Progression
Feature selection
Artificial intelligence

ABSTRACT

Background: Parkinson's disease (PD) is a chronic and progressive neurodegenerative disease with no cure, presenting a challenging diagnosis and management. However, despite a significant number of criteria and guidelines have been proposed to improve the diagnosis of PD and to determine the PD stage, the gold standard for diagnosis and symptoms monitoring of PD is still mainly based on clinical evaluation, which includes several subjective factors. The use of machine learning (ML) algorithms in spatial-temporal gait parameters is an interesting advance with easy interpretation and objective factors that may assist in PD diagnostic and follow up.
Research question: This article studies ML algorithms for: i) distinguish people with PD vs. matched-healthy individuals; and ii) to discriminate PD stages, based on selected spatial-temporal parameters, including variability and asymmetry.
Methods: Gait data acquired from 63 people with PD with different levels of PD motor symptoms severity, and 63 matched-control group individuals, during self-selected walking speed, was study in the experiments.
Results: In the PD diagnosis, a classification accuracy of 84.6 %, with a precision of 0.923 and a recall of 0.800, was achieved by the Naïve Bayes algorithm. We found four significant gait features in PD diagnosis: step length, velocity and width, and step width variability. As to the PD stage identification, the Random Forest outperformed the other studied ML algorithms, by reaching an Area Under the ROC curve of 0.786. We found two relevant gait features in identifying the PD stage: stride width variability and step double support time variability.
Significance: The results showed that the studied ML algorithms have potential both to PD diagnosis and stage identification by analysing gait parameters.

1. Introduction

An accurate diagnosis is required to differentiate Parkinson's disease (PD) from other neurological disorders and healthy individuals, and to improve PD treatment and follow up. Particularly, the diagnosis is challenging in the early stages of PD [1]. Despite a significant number of criteria and guidelines have been introduced to improve the PD diagnosis, the gold standard for PD diagnosis and symptoms monitoring is still based on clinical evaluation, which includes several subjective factors [2], and an accuracy from 75 % to 82 %, depending how perform it [3]. Also, the motor symptoms become more generalized with the disease progression - the trunk tends to bend forward, balance is reduced and gait becomes problematic [4] - which can hamper the identification of the PD stage.

Only the use of subjective clinical measures for PD evaluation can assure the validity and reliability of the PD diagnosis [5]. The Movement Disorder Society and previous studies have recently proposed other diagnostic aids to improve the accuracy. Machine learning (ML) based

Abbreviations: PD, Parkinson's disease; ML, machine learning; PwPD, people with Parkinson's disease; MMSE, Mini-Mental State Examination; H&Y, Hoehn and Yahr; UPDRS, Unified Parkinson's Disease Rating Scale.

* Correspondence to: São Paulo State University (Unesp), Department of Physical Education, Human Movement Research Laboratory (MOVI-LAB), Luiz Edmundo Carrijo Coube Avenue, 14-01, CEP: 17033-360, Bauru, Brazil.
E-mail address: fabio.barbieri@unesp.br (F.A. Barbieri).

https://doi.org/10.1016/j.gaitpost.2022.08.014
Received 8 February 2022; Received in revised form 16 August 2022; Accepted 18 August 2022
Available online 20 August 2022

approaches applied to gait parameters have been suggested to aid the diagnosis and follow up of PD [6,7]. Gait analysis is an interesting option with easy interpretation to be considered in PD diagnostic and follow up, mainly, due to gait impairments appearing early in PD and getting worse over time, differentiating the PD stages [8]. The spatial-temporal gait parameters are less prone to physiological parameters like age, height and weight [7], being a critical feature to identify the presence of PD and quantify the progression of the disease [6,7]. The spatial-temporal gait variability and asymmetry level are hallmarks of PD features [9,10]: while gait variability may predict falls in older adults and people with PD (PwPD), gait asymmetry may be helpful in early and moderate PD [6]. However, despite previous studies have shown interesting and relevant findings, there is not a consensus about the application of ML algorithms on gait parameters in order to diagnose and follow up PD.

As ML algorithms can extract useful patterns from processed data [5], they can potentially help physicians to diagnose PD, as well as quantify the disease progression. Although gait analysis by itself may not be a solution to PD, gait features could provide insights that are complementary to other sources of information [11]. Hence, from the spatial-temporal features of gait analysis, ML algorithms can find hidden patterns in the acquired data that can be important in PD diagnostic and follow up. Actually, by detecting changes in the gait pattern, ML algorithms were successfully used to identify PwPD from healthy subjects and to classify different PD stages [12,13]. Despite previous studies reported promising findings, indicating an accuracy between 73 % and 99.4 % to discriminate PD subjects from healthy subjects, or to classify different stages of PD [14,15], the literature has not identified the best ML based approach to help in PD diagnostic and follow up yet. Also, previous studies have included a limited sample size to perform ML techniques and the spatial-temporal gait variability and asymmetry level were not included in the ML model. Thus, our study seems to have a more comprehensive approach compared to others.

Therefore, to help defining such as ML based approach that can improve the PD diagnostic and follow up, the main purpose of this work was to study several ML algorithms to diagnose PD based on spatial-temporal gait characteristics. Specifically, our aims were to identify, based on ML algorithms, suitable spatial-temporal gait features that can better i) distinguish PwPD from age, height and body mass matched-healthy individuals, and ii) identify PD stages. The present study is a diagnostic accuracy observational study. The spatial-temporal gait parameters of PwPD and healthy individuals were used to produce an algorithm for the detection and classification of PD. The approach of this study could pave the basis for additional important perspectives for a precise diagnosis and stage determination in PD, determining gait biomarkers to help in the differential diagnosis. Also, the results of this study may help to predict disease progression, translating ML methods from the research field to clinical practice.

2. Materials and method

2.1. Participants

Sixty-three idiopathic PwPD and 63 healthy individuals from the community were recruited. Specifically, the PwPD were invited from Movement Disorders Clinics, such as ATIVA PARKINSON - UNESP - Bauru, Brazil. The PwPD were diagnosed by neurologists according to the UK Brain Bank Criteria. The exclusion criteria were: a) both groups: age below 50 years, score on the Mini-Mental State Examination (MMSE) of 24 points or below [16], and history of orthopaedic or vision problems that would make it impossible to perform the experimental protocol; b) PwPD: disease stage above 3 on Hoehn and Yahr (H&Y) scale [17], and starting PD medication lesser than three months before the data acquisition. Written informed consent was obtained from all participants according to the protocol approved by the local University Ethical Committee Board and it was conducted according to the Declaration of Helsinki.

To determine the motor condition and the disease stage, the PwPD were evaluated by a specialist using the motor part of the Unified Parkinson's Disease Rating Scale (UPDRS-III) [18] and H&Y scales, respectively. All evaluations were performed in the PD group approximately one hour after taking the dopaminergic medication (ON-state).

2.2. Gait task and analysis

The participants walked three times at their self-selected velocity for 8.5 m on an unobstructed walkway (1.5 m wide). The gait parameters were acquired using 10 Vicon Motion Systems® cameras (100 Hz). Two passive reflective markers were placed on each participants' foot (second metatarsal and calcaneus) in order to acquire the spatial-temporal gait parameters. The acquired data was filtered using a 5th order low-pass digital Butterworth filter (zero-lag) with a 6 Hz cut-off frequency.

The following spatial-temporal gait parameters were calculated: step/stride length, width, duration and velocity, single and double support time, and cadence. Additionally, step/stride variability and asymmetry (i.e., coefficient of variation and symmetry index, respectively) parameters were calculated. The average of three trials for each participant was used in the assessment of the studied ML algorithms and gait parameters.

2.3. Machine learning

2.3.1. Feature selection and classification

The implementation of the studied ML algorithms was done using the Python programming language on the Spyder integrated development environment (3.3.6 version). The used library packages were: scikit-learn, for the supervised ML algorithms; numpy and pandas, for data operations and mathematical computations; and matplotlib and seaborn, for graphs plotting.

In this study, we developed two different ML models for each case: first for PD diagnosis and another for the PD stage identification. We adopted the same strategy for each case: preprocessing step, feature selection, classification, and evaluation.

In the pre-processing step, the acquired data were randomly split into 80–20 %, in case of the PD diagnosis, as training and testing sets, respectively. In the second case, the split was stratified into 50–50 % to ensure that the ratio of the two classes was the same in both training and test sets, followed by feature scaling. Each PD stage was identified according to H&Y scale, mild PD corresponds to H&Y score from 1 to 2 and moderate PD matches with H&Y score from 2.5 to 3. The features were scaled in terms of median and quartiles, which is a technique useful when the data under study have many outliers [19].

Then, feature selection on the training set was performed in the same way in both objectives of the study. One form of feature selection is based on the univariate feature selection method, which considers each feature individually. So, a feature is discarded if it is only informative when combined with another feature. Choosing a test, for instance *f_classif*, and computing a threshold, all features with a too high p-value are discarded because they are unlikely to be related to the target class. *SelectKBest* is an example of a method which computes this threshold, by selecting a fixed *k* number of features [20].

In this work, the univariate feature selection method was used for selecting the important features; although, as a first step, features with a pairwise Pearson's correlation coefficient above 0.80 were removed. Following Evans [21] recommendations, 0.8–1.0 is considered very strong correlation between features, so they are not independent and should be removed. Because the univariate feature selection method [22] does not capture redundancy among features, correlation-based feature selection was applied before. However, the optimum number of features varied according to the used ML algorithm. Because of this, each classifier algorithm was used with its default parameters and the optimum number of features was chosen according to the corresponding highest achieved accuracy assessed by the 10-fold cross validation

method. Due to that, to identify PD stages, the accuracy was replaced by the Area Under the Curve (AUC) metric and a repeated 10-fold cross validation technique was adopted to evaluate the classification performance [23]. Furthermore, AUC is much suitable for imbalanced classification problems than accuracy [20], because a high accuracy is still possible achieved by an unskilful ML classification algorithm that only predicts the majority class [24].

The selected features from the training data set were used as inputs for the ML classification algorithms. In this work, in the two cases under study, we employed different ML algorithms: Naïve Bayes (NB) [25], Support Vector Machine (SVM) [26], Decision Tree (DT) [27], Random Forest (RF) [28], Logistic Regression (LR) [29], and Multilayer Perceptron (MLP) [30].

Before the learning process, a set of hyperparameters were determined by grid search. This technique can help to improve the performance of a ML algorithm by finding the optimal combination for the hyperparameter values. During this process, an exhaustive search is performed in order to specify different hyperparameter values [31].

As the dataset is composed of 63 individuals in PD stage identification, which is a reduced dataset, and consists of 42 mild and 21 moderate PwPD, the distribution of instances across the known classes is not equal, so the training set is slightly imbalanced. We reported the classification performance comparison between the five algorithms for the original training set and also for the balanced training set using the Synthetic Minority Oversampling Technique (SMOTE) [23] oversampling algorithm. This algorithm generates new synthetic samples by interpolation until minority class becomes of the same size as the majority class.

2.3.2. Classification evaluation

The performance of the studied ML classifiers on the test data was assessed using evaluation metrics such as accuracy, precision, recall, F1-score, and AUC value. Based on the outcome of a classification test, the number of true positives (TP) can be calculated as well as of true negatives (TN), false positives (FP) and false negatives (FN) [32]. Accuracy, which is one of the most used evaluation metrics, measures the proportion of correct predictions [33] over the total number of evaluated instances:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (1)$$

Precision is used to measure the positive patterns that are correctly predicted from the total predicted patterns in a positive class:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall represents the proportion of positive patterns that are correctly classified:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-score measures the harmonic mean between recall and precision values:

$$F1\text{-}score = \frac{2 * Recall * Precision}{Recall + Precision} \quad (4)$$

One of the popular ranking type metrics is AUC. Instead of the threshold and probability metrics, the AUC value reflects the overall ranking performance of a classifier. It was theoretically and empirically proven that the AUC is better than the Accuracy for evaluating the performance of a ML classifier.

3. Results

The demographic and clinical details of the participants are shown in Table 1. The PwPD and healthy individuals were matched by age, height, and body mass.

Table 1
Demographic and clinical characteristics (means ± standard deviations) of the studied neurologically healthy and PD individuals.

	Control Group	PD group	H&Y 1–2 (mild - n = 42)	H&Y 2.5–3 (moderate - n = 21)
M/F (n)	18M/45F	34M/29F	21M/21F	13M/8F
Age (yr)	68.0±6.0	68.7±7.5	68.0±7.5	70.4±7.5
Body mass (kg)	72.1±13.2	69.4±13.7	73.1±11.9	62.0±14.6
Height (m)	1.60±0.08	1.62±0.09	1.63±0.09	1.61±0.09
BMI (kg/m^2)	28.2±4.6	26.3±4.6	27.7±3.0	23.9±5.7
MMSE (pts)	27.9±2.0	27.4±2.0	27.7±2.0	27.1±1.9
Time from clinical diagnosis (yr)	–	5.33±3.67	5.2±3.6	5.5±3.8
UPDRS-III (pts)	–	27.4±10.6	24.7±10.1	32.9±9.7
H&Y (pts)	–	2.1±0.4	–	–

BMI: body mass index; MMSE: Mini – Mental State Examination; H&Y: Hoehn and Yahr; UPDRS-III: motor part of Unified Parkinson's Disease Rating Scale.

3.1. Performance in PD classification

In the first feature selection step, we calculated the Pearson correlation coefficients among 44 features. Fig. 1 shows the resulting correlation matrix after the highly correlated features have been removed (r>=0.8).

Fig. 2 shows the accuracy of the ML methods in correctly classifying healthy and PwPD. One can be observed that the NB algorithm significantly improved the classification accuracy from 61.5 % to 84.6 %. On the other hand, LR algorithm did not increase the classification accuracy after *SelectKBest*, remaining in 76.9 %. Furthermore, the RF algorithm did not change its accuracy (stayed in 57.7 %) after the highly correlated features have been removed, but improved it to 69.2 % after *SelectKBest*.

The NB algorithm with the optimum number of four features, which include step length, step velocity, step width and step width variability, provided the maximum accuracy (84.6 %), precision (0.923), recall (0.800), and F1-score (0.857). SVM with the RBF kernel achieved the higher AUC value (0.915) using the same features used with the NB algorithm (Table 2).

The optimized hyperparameters of the studied ML algorithms, which were found by grid search, are indicated in supplementary material 1 (S1). Fig. 3 shows the feature importance of the RF classifier in PD classification.

This classifier chose a total of twelve features and the three most important features were step length, step velocity variability and step velocity.

3.2. Performance in PD stage classification

The performance of the five studied algorithms on the original and balanced training sets is shown in Table 3. We used macro-averaging methods for evaluating the scoring metrics: precision, recall and F1-score. These methods give equal weight to all classes, no matter what their size is [20].

All algorithms showed a higher AUC value, standing out RF which was the best classifier with an AUC value of 0.786, followed by NB with an AUC of 0.771. In addition to the AUC value, the macro average (macro avg) of precision, recall and F1-score, are other important

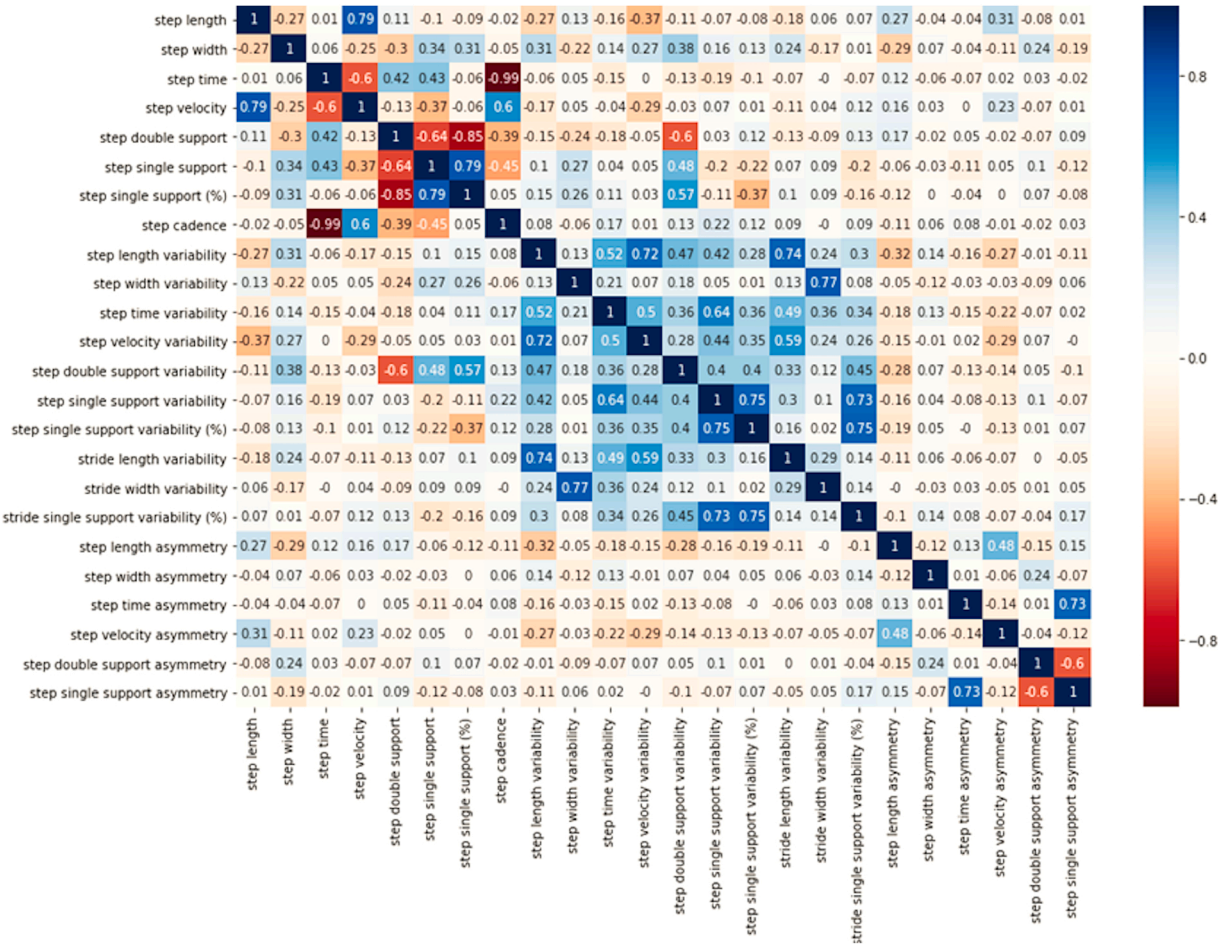

Fig. 1. Correlation matrix showing the Pearson correlation coefficients among the selected features.

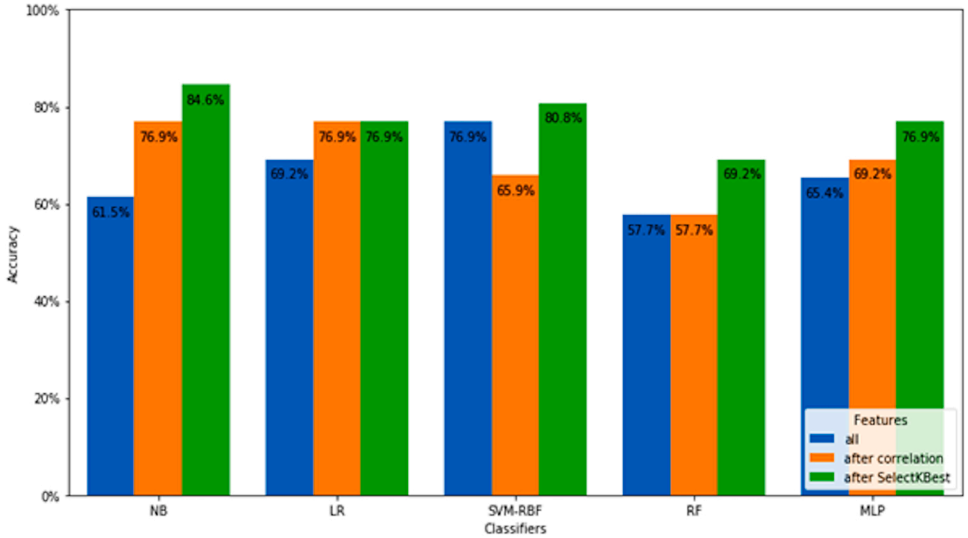

Fig. 2. Impact of the feature selection in the classification accuracy obtained by the studied ML algorithms. **NB**: Naïve Bayes; **LR**: Logistic Regression; **SVM**: Support Vector Machine, **RBF**: Radial Basis Function kernel, **RF**: Random Forest, **MLP**: Multilayer Perceptron.

measures to assess the quality of the classification results. The NB algorithm had achieved a higher precision macro avg, recall macro avg and F1-score macro avg.

4. Discussion

We studied the use of common ML algorithms to classify and rate PwPD using spatial-temporal gait parameters. We observed that as to the PD diagnosis, an accuracy of 84.6 % with a precision of 0.923 and a

Table 2
Performance values obtained by the studied ML algorithms trained on the found optimum number of features. These values are referred to distinguish PD group from control group in the training set.

Algorithms	Number of features	Accuracy (%)	Precision	Recall	F1-score	AUC
NB	4	84.6	0.923	0.800	0.857	0.891
LR	6	76.9	0.909	0.667	0.769	0.885
SVM-RBF	4	80.8	0.917	0.733	0.815	0.915
RF	12	69.2	0.889	0.533	0.667	0.739
MLP	6	76.9	0.909	0.667	0.769	0.885

NB: Naïve Bayes; LR: Logistic Regression; SVM: Support Vector Machine; RBD: Radial Basis Function; RF: Random Forest; MLP: Multilayer Perceptron.

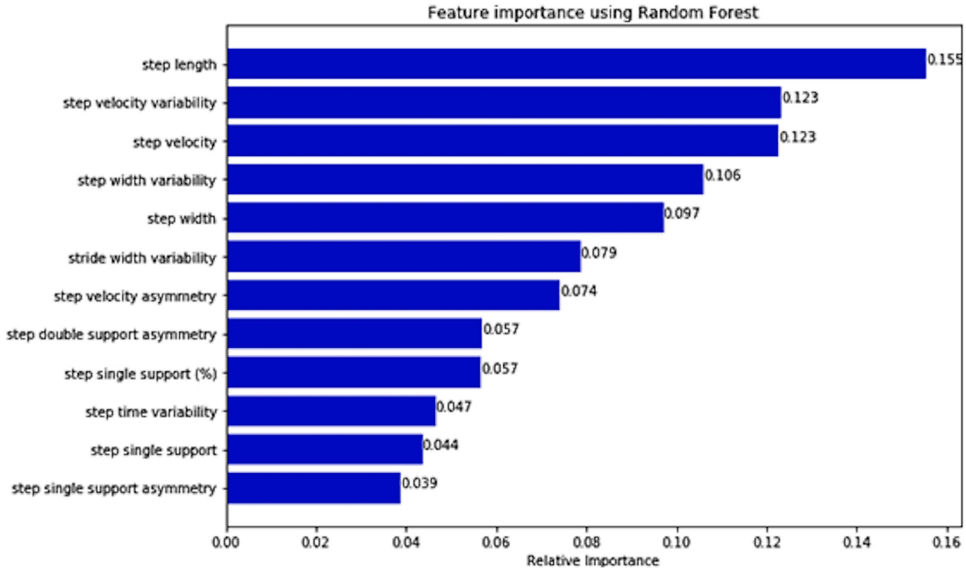

Fig. 3. Feature importance found as to the Random Forest classifier.

Table 3
Performance obtained by the different studied ML algorithms on the original training set and balanced training set with SMOTE.

Algorithms	Training set	AUC	Precision (macro avg)	Recall (macro avg)	F1-Score (macro avg)
NB	**Original**	0.766	0.737	0.680	0.691
	Balanced	0.771	0.729	0.745	0.733
LR	**Original**	0.762	0.328	0.500	0.396
	Balanced	0.766	0.692	0.699	0.695
SVM (linear)	**Original**	0.762	0.328	0.500	0.396
	Balanced	0.766	0.692	0.699	0.695
RF	**Original**	0.760	0.725	0.701	0.709
	Balanced	0.786	0.723	0.723	0.723
DT	**Original**	0.695	0.679	0.695	0.653
	Balanced	0.706	0.723	0.723	0.723

NB: Naïve Bayes; LR: Logistic Regression; SVM: Support Vector Machine; RF: Random Forest; DT: Decision Tree; AUC: Area Under the Curve; macro avg: macro average.

recall of 0.800 could be achieved by the NB algorithm. Regarding the gait parameters, we found four significant features in terms of PD diagnosis: step length, step velocity, step width and step width variability. In the second case, regarding the PD severity classification, RF reached an AUC value of 0.786 and NB achieved the best precision macro average (0.729), recall macro average (0.745), and F1-score macro average (0.733). We found two relevant features in rating the PD severity: stride width variability and step double support time variability.

The NB algorithm had the number of TP (PD cases) greater than the number of TN (healthy cases). In this classifier, the number of FN is always higher than the number of FP. In the medical context, the number of FN should be as low as possible, because it is a risk in a patient's treatment if the medical specialist uses the results of a classifier that predicts incorrectly PwPD as healthy (S2).

PD induces gait disturbances, such as reduced step length and reduced velocity [34], caused by bradykinesia, hypokinesia and rigidity. In addition, step width variability is related to active step-to-step adjustment by the central nervous system to maintain balance during gait [35]. Hence, step width was also included in our model, and unlike happened in previous studies, medio-lateral stability seems to be an important feature to classify PD. LR and MLP algorithms selected two more features than NB and SVM by including step time variability and step velocity asymmetry. The ML algorithms attributed less importance weight to the asymmetry parameters. Moreover, the RF algorithm also gave less importance to the asymmetry parameters.

Five different ML algorithms: NB, LR, SVM with RBF kernel, RF, and MLP, were applied in this study to spatial-temporal gait data. In a previous work, Rehman et al. [6] identified five clinical gait characteristics using a SVM classifier with a RBF kernel: mean step velocity, mean step length, step length variability, mean step width, and step width variability, and reached a classification accuracy between 73 % and 97 %, with 63–100 % of sensitivity and 79–94 % of specificity. Here, we found step length, step velocity, step width and step width variability as important parameters to classify PD. On the other hand, Wahid et al. [36] revealed significant differences in stride length, cadence, stance time, and double support time, and a RF classifier attained the best PD classification with an accuracy of 92.6 %. In Urcuqui et al. [37], the RF classifier also achieved higher accuracy (82 %) in comparison to other classifiers, and lower FN (23 %) and false positive (12 %) rates. They concluded that stride length and age were important features. Aich et al.

[38] extracted six spatial-temporal gait parameters, including step length, stride length, step time, stride time, step velocity, and stride velocity. Among all the ML techniques used in their study, RF detected the two states with an accuracy of 96.72 %, recall of 97.35 %, and precision of 96.92 %. In another study, Aich et al. [39] measured five important gait characteristics: step time, stride time, step length, stride length, and walking speed. The results showed that the DT classifier reached the best classification accuracy with a value equal to 88.46 %. Overall, previous studies identified step/stride length and step/stride time as important gait characteristics. Furthermore, RF and SVM classifiers obtained the best performance with an accuracy that varies between 72.67 % and 100 %. Some aspects may decrease performance evaluation metrics in our study compared to previous studies, including i) great dispersion in the gait parameters, including significant differences for sex, body mass and preferred gait speed. In addition, the inclusion of other gait parameters, such as muscle activity and kinetic parameters, should improve the precision of the algorithm, ii) despite H&Y is used as the main PD classification, the inclusion of full UPDRS could help to improve the performance evaluation metrics, and iii) reduced sample for training, increasing the risk of overfitting.

Neurodegenerative diseases, such as PD, increase the variability of walking and is related to falls. The center responsible for automatic gait control causes walking less automatic and more dependent on the cortical control [9]. Concerning the classification for different stages of PD, we used a total of five different ML algorithms: NB, LR, SVM linear, RF and DT. Each algorithm chose the optimum number of two features, which include stride width variability and step double support time variability. The selection of variability parameters as important characteristics to classify PD stage is interesting, because the variability of the temporal parameters of gait is related to mechanisms that regulate the movement rhythm and central pattern generator, and the variability of spatial parameters is associated to the balance control mechanisms.

Analysing Table 3, one can observe that class imbalanced had a negative impact, mainly on the LR and SVM algorithms' performance. Due to poor performance of these algorithms, we applied SMOTE based oversampling to solve the imbalance between the classes in order to improve the classification performance. The application of SMOTE technique is a novelty method for determining PD severity classification. Generally, the dataset is unbalanced due to the heterogeny of PD. To avoid errors in the calculation and classification, it is necessary to correct the dataset unbalance. Previous studies tried to solve this problem using cross-validation k-fold [12–15]. However, this technique did not correct the dataset unbalance, increasing the errors in the calculation.

In the current literature about PD, only a few number of studies are focused on the classification affected by different disease severity. Buongiorno et al. [40] used SVM and Artificial Neural Networks (ANN) classifiers for comparison. The results showed that the ANN classifier achieved the best accuracy by reaching a value of 95.0 % using six selected features for PD severity rating, but variability gait characteristics were not analysed. On the other hand, Balaji et al. [7] showed that the Decision Tree (DT) classifier attained an overall accuracy of 99.4 %, with the least negative predictive value of all classifiers, and the lowest misclassification rate for all stages considered: healthy, high, moderate, and mild.

We can conclude that spatial-temporal gait parameters together with ML techniques could support medical specialists in the identification and assessment of PwPD. The results showed that the studied ML algorithms have potential to diagnosis and follow up PD by analysing gait parameters. However, external validation is needed to confirm these results. In addition, ML techniques using spatial-temporal gait parameters in PD has clinical applications: i) it may support the diagnosis of PD by classifying people with PD with specific PD-related gait characteristics, ii) it has a significant impact on clinical practice by allowing identification of specific gait characteristics and further supporting tailored interventions focus on those specifics characteristics; iii) it increases the precision of PD identification and classification, making the clinical decisions more assertive.

As future work, we recommend exploring low-cost systems for data acquisition. A similar analysis can also be performed with wearable systems in the home environment. In addition to that, a larger number of individuals in the current dataset may improve the reliability of the ML algorithms; this can turn possible the use of more advanced ML algorithms such as deep learning, which require huge datasets for training. For enhancing the stage classification, we suggest assessing tremor data together with gait parameters. Also, in order to obtain a more complete diagnosis of PD, other motor and non-motor symptoms could be included in the ML classifiers.

CRediT authorship contribution statement

Marta Isabel A. S. N. Ferreira: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing, Visualization. **Fabio Augusto Barbieri**: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Vinícius Christianini Moreno**: Data curation, Writing – review & editing, Visualization. **Tiago Penedo**: Data curation, Formal analysis, Methodology, Investigation, Writing – review & editing, Visualization. **João Manuel R. S. Tavares**: Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Conflict of interest

All the authors declare no conflict of interest.

Acknowledgment

This study was financially supported by São Paulo Research Foundation (FAPESP) - grants: #14/20549-0, #17/19516-8, and #20/01250-4. FAB, VCM and TP would like to thank The National Council for Scientific and Technological Development (CNPq) for the personal funding support. In addtion, this study was financed in part by the Coordenaç ão de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.gaitpost.2022.08.014.

References

[1] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, G. Logroscino, Accuracy of clinical diagnosis of Parkinson disease, Neurology 86 (2016) 566 LP–566576, https://doi.org/10.1212/WNL.0000000000002350.
[2] L. di Biase, A. Di Santo, M.L. Caminiti, A. De Liso, S.A. Shah, L. Ricci, V. Di Lazzaro, Gait analysis in Parkinson's disease: an overview of the most accurate markers for diagnosis and symptoms monitoring, Sensors 20 (2020) 3529, https://doi.org/10.3390/s20123529.
[3] S. Caproni, C. Colosimo, Diagnosis and differential diagnosis of Parkinson disease, Clin. Geriatr. Med. 36 (2020) 13–24, https://doi.org/10.1016/j.cger.2019.09.014.
[4] H.H.P. Nguyen, M.A. Cenci. Behavioral Neurobiology of Huntington's Disease and Parkinson's Disease, 1st ed., Springer Berlin Heidelberg, Berlin, Heidelberg, 2015 https://doi.org/10.1007/978-3-662-46344-4.
[5] N. Kour, Sunanda, S. Arora, Computer-vision based diagnosis of Parkinson's disease via gait: a survey, IEEE Access 7 (2019) 156620–156645, https://doi.org/10.1109/ACCESS.2019.2949744.
[6] R.Z.U. Rehman, S. Del Din, Y. Guan, A.J. Yarnall, J.Q. Shi, L. Rochester, Selecting clinically relevant gait characteristics for classification of early parkinson's disease: A comprehensive machine learning approach, Sci. Rep. 9 (2019) 1–12, https://doi.org/10.1038/s41598-019-53656-7.
[7] E. Balaji, D. Brindha, R. Balakrishnan, Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease, Appl. Soft Comput. 94 (2020), 106494, https://doi.org/10.1016/j.asoc.2020.106494.
[8] L. Rochester, D. Jones, V. Hetherington, A. Nieuwboer, A.-M. Willems, G. Kwakkel, E. Van Wegen, Gait and gait-related activities and fatigue in Parkinson's disease:

what is the relationship? Disabil. Rehabil. 28 (2006) 1365–1371, https://doi.org/10.1080/09638280600638034.
[9] L. Simieli, L.T.B. Gobbi, D. Orcioli-Silva, V.S. Beretta, P.C.R. Santos, A.M. Baptista, F.A. Barbieri, The variability of the steps preceding obstacle avoidance (approach phase) is dependent on the height of the obstacle in people with Parkinson's disease, PLoS One 12 (2017), e0184134, https://doi.org/10.1371/journal.pone.0184134.
[10] F.A. Barbieri, L. Simieli, D. Orcioli-Silva, A.M. Baptista, M. Borkowske Pestana, V. Spiandor Beretta, P.C.R. dos Santos, L.T.Bucken Gobbi, Obstacle avoidance increases asymmetry of crossing step in individuals with Parkinson's disease and neurologically healthy individuals, J. Mot. Behav. 50 (2018) 17–25, https://doi.org/10.1080/00222895.2016.1271303.
[11] E. Rastegari, S. Azizian, H. Ali, Machine learning and similarity network approaches to support automatic classification of parkinson's diseases using accelerometer-based gait analysis, in: Proc. 52nd Hawaii Int. Conf. Syst. Sci., Hawaii, 2019: pp. 4231–4242. doi:10.24251/HICSS.2019.511.
[12] C. Ricciardi, M. Amboni, C. De Santis, G. Ricciardelli, G. Improta, G. D'Addio, S. Cuoco, M. Picillo, P. Barone, M. Cesarelli, Machine learning can detect the presence of Mild cognitive impairment in patients affected by Parkinson's Disease, in: 2020 IEEE Int. Symp. Med. Meas. Appl., IEEE, 2020: pp. 1–6. doi:10.1109/MeMeA49120.2020.9137301.
[13] C. Caramia, D. Torricelli, M. Schmid, A. Muñoz-Gonzalez, J. Gonzalez-Vargas, F. Grandas, J.L. Pons, IMU-based classification of Parkinson's disease from gait: a sensitivity analysis on sensor location and feature selection, IEEE J. Biomed. Heal. Inform. 22 (2018) 1765–1774, https://doi.org/10.1109/JBHI.2018.2865218.
[14] Y. Mittra, V. Rustagi, Classification of Subjects with Parkinson's Disease Using Gait Data Analysis, in: 2018 Int. Conf. Autom. Comput. Eng., IEEE, Greater Noida, India, 2018: pp. 84–89. doi:10.1109/ICACE.2018.8687022.
[15] S. Shetty, Y.S. Rao, SVM based machine learning approach to identify Parkinson's disease using gait analysis, in: 2016 Int. Conf. Inven. Comput. Technol., IEEE, 2016: pp. 1–5. doi:10.1109/INVENTIVE.2016.7824836.
[16] O.P. Almeida, Mini mental state examination and the diagnosis of dementia in Brazil, Arq. Neuropsiquiatr. 56 (1998) 605–612, https://doi.org/10.1590/S0004-282×1998000400014.
[17] M.M. Hoehn, M.D. Yahr, Parkinsonism: onset, progression, and mortality, 427–427, Neurology 17 (1967), https://doi.org/10.1212/WNL.17.5.427.
[18] S.R.L.E. Fahn, Unified Parkinson's disease rating scale, Recent Dev. Park. Dis. (1987).
[19] A. Giussani, Applied Machine Learning with Python, 1st ed, Egea, Milano, 2020.
[20] A.C. Müller, S. Guido. Introduction to Machine Learning with Python: A Guide for Data Scientists, 1st ed.,, O'Reilly Media,, Sebastopol, CA, 2016.
[21] J.D. Evans, Straightforward Statistics for The Behavioral Sciences, Brooks/Cole Pub, Co., Pacific Grove, 1996.
[22] V. Jain, J.M. Chatterjee, Machine Learning with Health Care Perspective: Machine Learning and Healthcare, 1st ed., Springer International Publishing, Cham, 2020.
[23] T. Amr, Hands-On Machine Learning with scikit-learn and Scientific Python Toolkits: A practical guide to implementing supervised and unsupervised machine learning algorithms in Python, 1st ed, Packt Publishing, Birmingham, UK, 2020.
[24] J. Brownlee, Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning, 1st ed, Machine Learning Mastery, Calle de San Francisco, 2020.
[25] M. Granik V. Mesyura, Fake news detection using naive Bayes classifier, in: 2017 IEEE First Ukr. Conf. Electr. Comput. Eng., IEEE, Kiev, 2017: pp. 900–903. doi:10.1109/UKRCON.2017.8100379.
[26] M.M. Rahman, Y. Ghasemi, E. Suley, Y. Zhou, S. Wang, J. Rogers, Machine learning based computer aided diagnosis of breast cancer utilizing anthropometric and clinical features, IRBM (2020), https://doi.org/10.1016/j.irbm.2020.05.005.
[27] N. Pentreath, Machine Learning with Spark, 1st ed, Packt Publishing, BIRMINGHAM, MUMBAI, 2015.
[28] G. Rebala, A. Ravi, S. Churiwala, An Introduction to Machine Learning, 1st ed, Springer International Publishing, Cham, 2019.
[29] A.K. Dwivedi, Performance evaluation of different machine learning techniques for prediction of heart disease, Neural Comput. Appl. 29 (2018) 685–693, https://doi.org/10.1007/s00521-016-2604-1.
[30] G. Solana-Lavalle, J.-C. Galán-Hernández, R. Rosas-Romero, Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features, Biocybern. Biomed. Eng. 40 (2020) 505–516, https://doi.org/10.1016/j.bbe.2020.01.003.
[31] S. Raschka, Python Machine Learning, 1st ed, Packt Publishing, Birmingham, UK, 2015.
[32] A. Dumortier, E. Beckjord, S. Shiffman, E. Sejdic, Classifying smoking urges via machine learning, Comput. Methods Prog. Biomed. 137 (2016) 203–213, https://doi.org/10.1016/j.cmpb.2016.09.016.
[33] M. Hossin, M.N. Sulaiman, A review on evaluation metrics for data classification evaluations, Int. J. Data Min. Knowl. Manag. Process. 5 (2015) 1–11, https://doi.org/10.5121/ijdkp.2015.5201.
[34] R. Pahwa, K.E. Lyons. Handbook of Parkinson's Disease, 5th ed.,, Taylor & Francis, Boca Raton, FL, 2013.
[35] D.S. Peterson, F.B. Horak, Neural control of walking in people with Parkinsonism, Physiology 31 (2016) 95–107, https://doi.org/10.1152/physiol.00034.2015.
[36] F. Wahid, R.K. Begg, C.J. Hass, S. Halgamuge, D.C. Ackland, Classification of Parkinson's disease gait using spatial-temporal gait features, IEEE J. Biomed. Heal. Inform. 19 (2015) 1794–1802, https://doi.org/10.1109/JBHI.2015.2450232.
[37] C. Urcuqui, Y. Castaño, J. Delgado, A. Navarro, J. Diaz, B. Muñoz, J. Orozco, Exploring Machine Learning to Analyze Parkinson's Disease Patients, in: 2018 14th Int. Conf. Semant. Knowl. Grids, IEEE, Guangzhou, China, 2018: pp. 160–166. doi:10.1109/SKG.2018.00029.
[38] S. Aich, J. Youn, S. Chakraborty, P.M. Pradhan, J.-H. Park, S. Park, J. Park, A. Supervised, Machine learning approach to detect the on/off state in Parkinson's disease using wearable based gait signals, Diagn. (Basel, Switz.) 10 (2020) 1–18, https://doi.org/10.3390/diagnostics10060421.
[39] S. Aich, P.M. Pradhan, S. Chakraborty, H.-C. Kim, H.-T. Kim, H.-G. Lee, I.H. Kim, M. Joo, S. Jong Seong, J. Park, Design of a machine learning-assisted wearable accelerometer-based automated system for studying the effect of dopaminergic medicine on gait characteristics of Parkinson's patients, J. Healthc. Eng. 2020 (2020) 1–11, https://doi.org/10.1155/2020/1823268.
[40] D. Buongiorno, I. Bortone, G.D. Cascarano, G.F. Trotta, A. Brunetti, V. Bevilacqua, A low-cost vision system based on the analysis of motor features for recognition and severity rating of Parkinson's Disease, BMC Med. Inform. Decis. Mak. 19 (2019) 1–13, https://doi.org/10.1186/s12911-019-0987-5.