Computational and Structural Biotechnology Journal 27 (2025) 778–784

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Testing machine learning algorithms to evaluate fluctuating and cognitive profiles in Parkinson's disease by motion sensors and EEG data

Giovanni Mostile [a,b,1], Salvatore Quattropani [c,d,1], Federico Contrafatto [a], Claudio Terravecchia [a], Michelangelo Riccardo Caci [a], Alessandra Chiara [a], Calogero Edoardo Cicero [a], Giulia Donzuso [a], Alessandra Nicoletti [a], Mario Zappia [a,*]

[a] Department of Medical, Surgical Sciences and Advanced Technologies "G.F. Ingrassia" (DGFI), University of Catania, Catania, Italy
[b] OASI Research Institute - IRCCS, Troina, Italy
[c] Department of Electrical, Electronics and Computer Engineering (DIEEI), University of Catania, Catania, Italy
[d] National Inter-University Consortium for Telecommunications (CNIT), Catania, Italy

ARTICLE INFO

Keywords:
Machine learning
Wereable sensors
EEG
Parkinson
Levodopa
Cognitive impairment

ABSTRACT

Objective: We aimed to test machine learning algorithms for classifying fluctuating and cognitive profiles in Parkinson's Disease (PD) by using multimodal instrumental data.
Methods: Data of motion transducers while performing instrumented Timed-Up-and-Go test (iTUG) (N = 30 subjects) and EEG (N = 49 subjects) from PD patients were collected. Study patients were classified based on cognitive profile ("mild cognitive impairment" by standardized criteria vs "normal cognition") and L-dopa acute motor response ("fluctuating" vs "stable") to be analyzed by machine learning algorithms and compared with historical control data from healthy subjects group-matched by age for both iTUG and EEG study (for iTUG: N = 31 subjects; for EEG: N = 27 subjects).
Results: Artificial Neural Network-based models revealed the best performances when applied to specific phases of the iTUG in differentiating PD vs controls (91 % accuracy) as well as in differentiating cognitive profile (95 % accuracy) and motor response status (96 % accuracy) among PD subjects. K-Nearest Neighbors revealed best performances when applied to EEG data in discriminating PD vs controls (85 % accuracy). Random Forest Classifier revealed best performances when applied to EEG data in differentiating cognitive profile (96 % accuracy) and motor response status (91 % accuracy) among PD subjects.
Conclusions: By processing multimodal instrumental data, specific machine learning algorithms have been identified which discriminated L-dopa responsiveness and cognitive profile in PD. Further studies are needed to validate them in independent samples using a user-friendly software interface created *ad hoc*.

1. Background

Parkinson's Disease (PD) is a widely heterogeneous condition in terms of clinical presentation, disease progression and treatment response [1]. The clinical hallmark of PD is a motor syndrome encompassing bradykinesia, rest tremor and rigidity resulting from nigral degeneration and striatal dopaminergic depletion [1]. Moreover, postural and gait disorders characterized by reduced speed and step length, impaired rhythmicity as well as turning difficulties usually develop during the disease course, representing a major burden affecting the independence and quality of life of PD patients and their caregivers [2]. Beside motor features, a variety of non-motor symptoms encompassing autonomic, sleep, neuropsychiatric and cognitive disorders also characterize PD as a consequence of neurodegeneration involving extranigral non-dopaminergic structures [1]. In particular, cognitive dysfunction represents one of the most common non-motor symptoms in PD, ranging from mild cognitive impairment to dementia, with further negative implications for functioning, caregiver burden, quality of life and health-related costs [3].

Notably, a strict interplay between cognition and gait performances has been described in PD, supported by clinical, functional and morphological evidences showing an involvement of several brain

* Corresponding author.
E-mail address: m.zappia@unict.it (M. Zappia).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2025.02.019
Received 15 November 2024; Received in revised form 18 February 2025; Accepted 18 February 2025
Available online 19 February 2025

cortical (i.e. prefrontal cortex) and subcortical regions (i.e. pedunculopontine nucleus) other than basal ganglia both in PD-related gait disorders and cognitive decline [2].

According to current diagnostic criteria, PD diagnosis primarily relies on clinical assessment, in which a clear beneficial response to dopaminergic treatment represents a key supportive diagnostic point [4]. However, a great inter-individual variability in dopaminergic response has been described since it represents a very complex phenomenon resulting from a combination of treatment-responsive, treatment-resistant as well as treatment-induced worsening effects often coexisting in the same patient [5,6]. Notably, variable effects of dopaminergic treatments have been demonstrated on gait disturbances in PD, reflecting the complexity of their underlying pathophysiology [2].

In the last years, a growing attention has been focused on objective instrumental biomarkers as possible tools in PD characterization. In particular, motion sensors-based gait analysis has been demonstrated as a reliable approach to identify subclinical gait disorders also in early stages of the disease [7], for differential diagnosis [8] as well as in monitoring the treatment effects [9]. Similarly, electrocortical activity (EEG) analysis provided quantitative insights as a biomarker in PD-related cognitive impairment [10] as well as a possible tool to investigate the relationship between cognition and gait disorders in this condition [11].

Data-driven analysis of clinical and instrumental data from monitoring systems has been extensively used in clinical research for differentiating and phenotyping pathological conditions, including PD. In particular, different machine learning techniques have been used to differentiate PD patients from controls, principally focusing on gait and mobility data, including supervised [12–14] and unsupervised [15] algorithms. To our knowledge, there are no consistent results on literature concerning the use of machine learning algorithms applied to multimodal data from different instrumental sources, specifically motion sensors and EEG, to identify fluctuating motor response and early-stage cognitive deficits in PD subjects. The objective of the study was to test these algorithms as well as to develop a software for their external validation on independent samples.

2. Methods

2.1. Data sampling

2.1.1. Study patients selection and evaluation

Clinical and instrumental data from independent cohorts of PD patients attending the Neurology Clinic in Catania, Italy were retrospectively selected to be analyzed. All study patients satisfied diagnostic criteria for clinically definite PD [4].

A first cohort included PD patients who performed instrumented Timed-Up-and- Go test (iTUG) using motion transducers (N = 30 subjects) [8,16,17]. A second cohort included PD patients who performed a standardized EEG study (N = 49 subjects) [10,18,19]. We excluded all patients who received a different diagnosis during follow-up visits and those who satisfied the diagnostic criteria for dementia [10].

Study subjects underwent a comprehensive neurological examination performed by movement disorder specialists. PD severity was evaluated in accordance with the Hoehn-Yahr stage and Unified Parkinson Disease Rating Scale - Motor Examination (UPDRS-ME). Cumulative daily dosage of dopaminergic drugs was converted and provided using the Levodopa Equivalent Dosage (LED) [20].

Evaluations for all subjects in both cohorts included a comprehensive neuropsychological assessment and a standardized evaluation of clinical response to L-dopa by short-term evaluation [16,18]. Diagnosis of PD-Mild Cognitive Impairment (PD-MCI) was made according to the Movement Disorder Society Task Force Level II criteria [10,21]. Otherwise, patients were defined as PD with Normal Cognition (PD-NC).

Patient's clinical evaluations were performed during their own "practical-off" motor state or "off" state, thus before taking the first daily dose of the dopaminergic drug after an overnight fast, as well as during their "on" state at the peak of dose after assuming the first daily dose of the dopaminergic medication. Based on a clinically detectable motor response on UPDRS-ME after the single oral dose of L-dopa, we classified patients in "fluctuating" vs "stable" setting a percent clinically-detectable motor response cut-off equal to 20.

We finally compared patients data with historical control data obtained from healthy subjects group-matched by age for both iTUG and EEG study (for iTUG: N = 31 subjects; for EEG: N = 27 subjects).

A graphical explanation of the adopted study methods is provided in Fig. 1.

2.1.2. iTUG test

Patients performed the TUG test wearing the inertial sensor BTS G-WALK (BTS Bioengineering S.p.A., Italy) on a waist belt covering the L4–L5 inter-vertebral space. The portable system consists in a wireless network of inertial sensors for human movement analysis. The sensors are controlled by a data logger unit. Each sensor is sized 62 mm × 36 mm × 16 mm, weighs 60 g. The unit is composed of a three-axis accelerometer (max range ± 6 g), a three-axis gyroscope (full scale ± 300 deg/s), and a three-axis magnetometer (full scale ± 6 gauss). The system is connected to a computer via Bluetooth. At the end of the measurement, data are automatically processed by a dedicated software (BTS G-STUDIO), which automatically provides information about fluency of sitting and rising movements, acceleration, speed and angle analysis during turning and walking pattern analysis.

Subjects were instructed to perform the task in "off" state. We used route with standardized length, asking patients to perform task at their own walking speed. The system differentiates the test in six phases: (a) sit-to-stand phase, (b) forward phase; (c) mid-turning phase; (d) backward phase; (e) turning-before-sitting phase; (f) stand- to-sit phase.

The following parameters were recorded during the different test phases: duration of phases, accelerations in antero-posterior (AP), medio-lateral (ML) and vertical axes, average and peak angular speeds during turning.

Algorithm for the automatic segmentation of iTUG into the six phases and the extraction of the related measures have been tested in previously published works [8,16,17].

2.1.3. EEG recordings

Resting (eye-closed task) EEG recording was performed on treatment using 19 electrodes placed according to the international 10–20 International System (Fp2, F4, C4, P4, O2, F8, T4, T6, Fz, Cz, Pz, Fp1, F3, C3, P3, O1, F7, T3, T5), with unipolar derivation and common reference (G2) located between Fz and Cz (acquisition system: SystemPLUS ver.1.02.1109 for BRAIN-QUICK BQ132 S, Micromed). The impedance of the EEG signal was kept between 2 and 10 kohm at the homologous sites. The signal underwent notch filter set at 50 Hz to remove artefacts due to AC current field and bandwidth filtering (1.6 and 30 Hz), being digitized using the default sampling rate.

Five visually selected epochs of 4-s artefacts-free EEG signal of clear-cut wakefulness, excluding the EEG activity of drowsiness, for each patient were randomly selected by an expert electroencephalographist to ensure quality of samples [10,18,19]. The adopted methodology is based on previous standardized protocols. We used short signal epochs with the assumption they may preserve intrinsic characteristics of signal as longer recordings based on long-range temporal correlations and scaling behavior demonstrated in human EEG [19–22].

2.2. Pre-processing and features extraction

In the realm of Artificial Intelligence (AI), particularly in the domain of biometric parameters, the extraction of relevant features plays a pivotal role in enhancing the efficiency and accuracy of predictive models. This chapter delves into the rationale behind the necessity of feature extraction during preprocessing stages in AI systems focused on

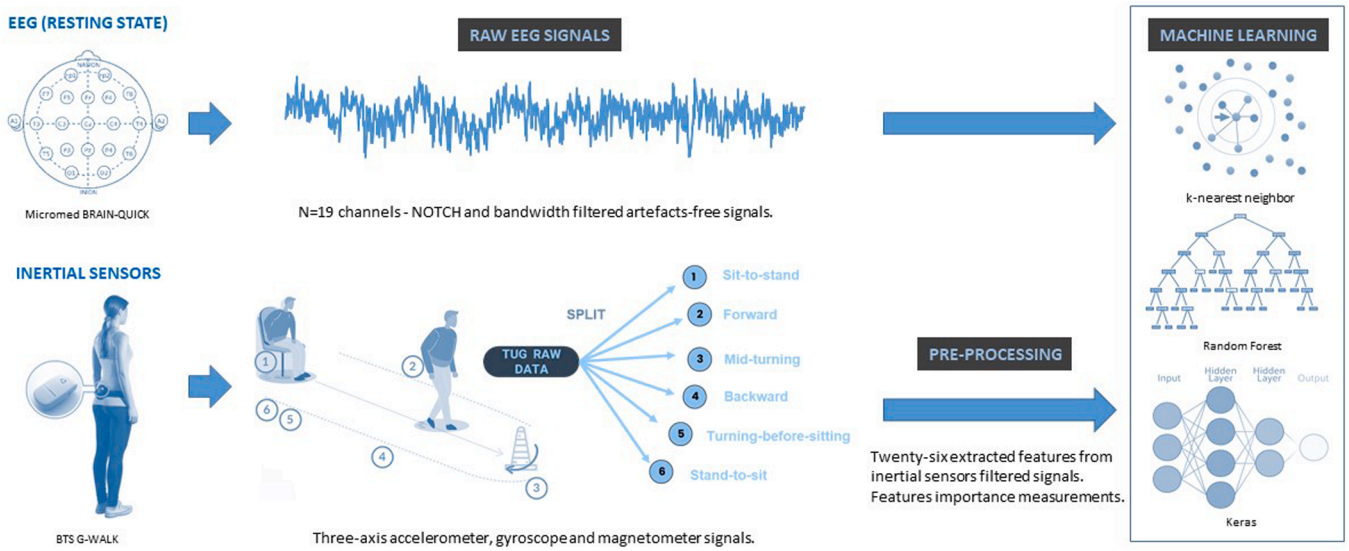

Fig. 1. Graphical explanation of data extraction, processing and analysis.

biometric parameters.

2.2.1. *iTUG signals*

Pre-processing operations were preliminary performed to improve data quality and remove any unwanted noise. These operations included signal filtering to reduce noise, normalization techniques to make data comparable and outliers' identification and managing. Pre-processing analysis performed on iTUG data and feature extraction details have been provided on Supplemental Material S1.

Twenty-six extracted features from the iTUG inertial data were used as input for classification models and for data analysis in order to discriminate PD from healthy controls using controls' historical data group-matched by age, as well as to differentiate PD-MCI vs PD-NC or PD "fluctuating" vs "stable" PD to L-dopa therapy. Processed results were saved in data frames along with the relevant variable labels using binary values (i.e. 1 for PD, 0 for controls).

2.2.2. *EEG signal*

For EEG signals, no functions were applied for data processing or feature extraction. Raw data recorded by the electrodes were kept in order to preserve the full informational content of the brain signals and leverage the capability of advanced machine learning algorithms to autonomously extract relevant latent features. The use of minimally preprocessed data may ensure that no potentially valuable information is discarded prematurely. Advanced machine learning models, particularly those employing non-linear architectures, are adept at handling complex, high-dimensional datasets and can discern relevant patterns while minimizing overfitting to noise or irrelevant features. This approach is supported by recent studies demonstrating that deep learning methods, in particular, perform well on raw EEG data without relying on extensive preprocessing, as preprocessing can inadvertently introduce biases or reduce the variability in the data that may contain valuable information [23]. The resulting datasets, composed of EEG data combined from multiple.csv files for each patient, were used for analysis and modeling of biometric data using machine learning algorithms.

Specifically, The EEG data was structured in a long format, where each.csv file contained rows corresponding to epochs and columns representing EEG channels. Data was then aggregated at the subject level: all epochs and channels were combined to create a single feature vector per subject. This ensured the classification task focused on distinguishing subjects rather than individual epochs. By aggregating the data, we avoided violating the independence of observations assumption, aligning with best practices for subject-level classification as discussed in the referenced article [24].

As for iTUG traces, data were used as input for classification models and for data analysis in order to discriminate PD from healthy controls using controls' historical data group-matched by age, as well as to differentiate PD-MCI vs PD-NC or "fluctuating" vs "stable" PD to L-dopa therapy as binary labels.

2.2.3. *Data Analysis*

After performing feature extraction, data were analyzed before feeding them into the artificial intelligence models. This step is crucial as it allows for a deeper understanding of the dataset and facilitates the identification of underlying patterns and relationships. Details have been provided in Supplemental Material S2. Briefly, feature importance (as shown in Figs. S2–2 and reported in Tables S2-4, S2–5, S2–6) was determined using different methods depending on the model. For tree-based models like Random Forest, feature importance was calculated based on the mean decrease in impurity, or the Gini index, reflecting how much each feature contributes to reducing uncertainty in the model. Additionally, permutation importance was applied to assess the impact of each feature by measuring the drop in model performance when the feature values were randomly shuffled. A feature importance ranking from 0 to 10 was also derived by normalizing the importance values, with higher ranks assigned to features that had a larger impact on model accuracy and decision-making. Principal Component Analysis (PCA) was used to visualize the distribution of data points in reduced-dimensional spaces and identify any clustering patterns (Figure S2-3). Correlation matrices were used to examine the relationships between different features and identify potential collinearities (Figure S2-4).

2.3. *Machine learning algorithms*

Performances of the following supervised algorithms were evaluated using the acquired dataset. To ensure a robust evaluation of model performance and minimize any potential bias from a single train-test split, we performed a 10-fold cross-validation for all adopted machine learning algorithms. The dataset was split into 10 subsets, and the model was trained on 9 subsets while the remaining subset was used for testing.

First, the K-Nearest Neighbors (KNN) algorithm was used. The algorithm was trained on each training subset, and the trained model was used to predict labels for the corresponding test subset. We calculated the confusion matrix for each fold and averaged the performance metrics, including accuracy, precision, recall, and F1-score, to obtain a more reliable esteem of model performance. These averaged metrics were

reported to evaluate the overall performance of the KNN algorithm.

A second analysis using the Random Forest Classifier (RFC) algorithm was performed. The RFC was built with 100 decision trees and a random seed of 42. The model was trained and tested in each fold, and performance metrics were averaged across all folds. The final reported metrics represent the mean of the values obtained from the cross-validation process, ensuring a more generalizable evaluation of the model.

Finally, a Python code was used to implement an Artificial Neural Network (ANN) for classification using the Keras library with a TensorFlow backend. The class labels of the dataset were converted into a one-hot encoding format. The sequential artificial neural network consisted of 3 layers: the first with 64 neurons, the second with 32 neurons, and the third layer with a number of neurons equal to the number of classes. The model was trained for 90 epochs with a batch size of 64. Accuracy, precision, recall, and F1-score were calculated for each fold, and the final performance metrics were the average of those obtained from the cross-validation procedure.

2.4. Models validation

For each phase of the iTUG analysis and for each type of conducted analysis, multiple models were evaluated for accuracy. Specifically, based on the iTUG phases, six models were selected for PD vs healthy controls classification, six models for PD-MCI vs PD-NC, and six models for the L-dopa response adopted classification. Additionally, for EEG analysis, three additional models were chosen for PD vs healthy controls classification, PD-MCI vs PD-NC, and L-dopa response classification. For each model, we used a 10-fold cross-validation strategy to evaluate their performance and reduce potential bias. Furthermore, in order to enhance model performance, we fine-tuned specific settings and configurations of each model (such as the number of trees in the Random Forest or the number of neighbors in KNN) based on validation results. After optimization, the models were re-evaluated and their final performance metrics were averaged across the cross-validation folds to provide a reliable estimate of generalization ability.

2.5. Interface development and automated output

An executable platform with a graphical interface was then developed integrating the tested algorithms (see Fig. 2). It was created with the aim to provided accuracy outputs on iTUG and EEG independent data on tested classification outputs even for external validation of the developed machine learning algorithms (GitHub link to the Python-based GUI code has been provided: https://github.com/SalvoQuattropani/ParkinsonAI-Toolkit.git).

3. Results

3.1. Clinical characteristics

Clinical characteristics of the two PD cohorts used for instrumental data acquisition are summarized in Table 1. We used as reference historical data from healthy controls group-matched by age for both iTUG and EEG study (for iTUG: N = 31 subjects; for EEG: N = 27 subjects).

3.2. Machine learning algorithms performances based on iTUG data: Training and validation

In total, 390 acquisitions were used for the iTUG analysis and 245 for the EEG data. For the EEG data, 150 acquisitions were allocated for training, 50 for validation, and 45 for the test set. For the iTUG data, 255 acquisitions were used for training, 85 for validation, and 50 for the test set.

The KNN classification algorithm, when trained to discriminate PD vs Controls, did not manage to discriminate classes correctly in the inference phase although presented a percent accuracy of 91 during the training phase.

When trained to discriminate PD-MCI vs PD-NC, the algorithm failed to discriminate classes correctly in the inference phase although it

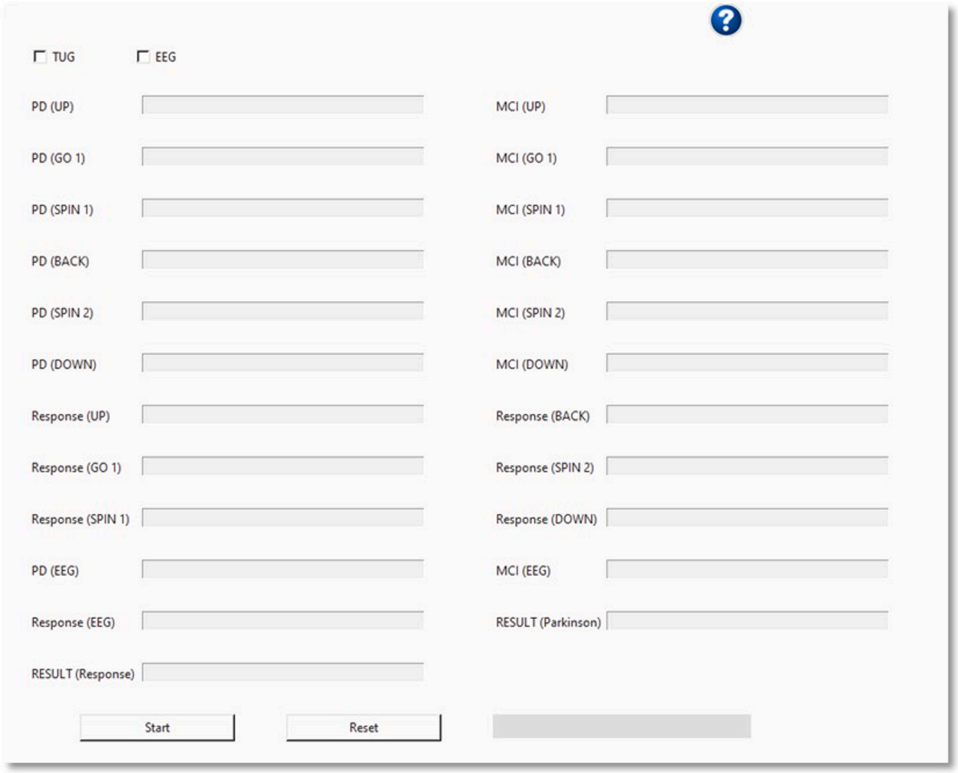

Fig. 2. Interface solution for automated classification of iTUG and EEG data.

Table 1
Clinical characteristics of study PD patients of the two data cohorts.

	All patients	PD-NC	PD-MCI	Fluctuating	Stable
iTUG Cohort (A)	**N = 30**	**N = 20**	**N = 10**	**N = 18**	**N = 12**
Sex (no. females)	11	7	4	7	4
Age (years)	62.4 ± 10.3	60.3 ± 9	66.8 ± 10.4	62.9 ± 10.2	62.6 ± 10.8
Disease duration (years)	4 ± 3.2	4 ± 3.2	5 ± 3.2	4 ± 3.2	4 ± 2.6
Age at onset (year)	59 ± 10.4	59 ± 10	59 ± 10.6	59 ± 10	59 ± 10.9
UPDRS-ME score (baseline)	29 ± 11.4	29 ± 10.8	31 ± 11.2	29 ± 10.8	30 ± 10.8
Hoehn-Yahr stage	2 ± 0.4	2 ± 0.4	2 ± 0.4	2 ± 0.4	2 ± 0.4
LED (mg/die)	322 ± 162	323 ± 165.4	336 ± 169.2	323 ± 165.4	319 ± 152.4
EEG Cohort (B)	**N = 49**	**N = 33**	**N = 16**	**N = 36**	**N = 13**
Sex (no. females)	21	15	6	17	4
Age (years)	62.6 ± 8.8	60.6 ± 9	66.6 ± 8.7	61.7 ± 8.8	64.8 ± 9
Disease duration (years)	4.6 ± 4	4.4 ± 4	4.6 ± 4	4.6 ± 4	4.7 ± 4.1
Age at onset (years)	57.5 ± 9	57.7 ± 9.3	57.6 ± 9	57.5 ± 9	58.1 ± 9.4
UPDRS-ME score (baseline)	35 ± 16.3	35 ± 16.6	35.5 ± 16.5	35 ± 16.3	37 ± 16.7
Hoehn-Yahr stage	2 ± 0.5	2 ± 0.5	2 ± 0.5	2 ± 0.5	2 ± 0.5
LED (mg/die)	409.7 ± 576.2	405.2 ± 587.7	417.7 ± 586.5	409.7 ± 576.2	434.9 ± 615.3

Legend: Data are frequencies and mean, s.d. values. PD-NC (PD with Normal Cognition); PD-MCI (PD with Mild Cognitive Impairment). Reference: historical data from healthy controls group-matched by age for both iTUG and EEG study (for iTUG: N = 31 subjects; for EEG: N = 27 subjects).

presented a percent accuracy of 93 during the training phase.

When trained to discriminate Fluctuating vs Stable response to therapy, it presented a percent accuracy of 60 during the training phase; in inference phase, it managed to correctly discriminate the classes. In particular, it demonstrated better performance for the phases of Sit-to-stand (88 percent of accuracy) and Forward (91 percent of accuracy).

The RFC algorithm, when trained to discriminate PD vs Controls, presented a percent accuracy ranging from 87 to of 91 percent across various iTUG phases during training, proving to be very efficient in the inference phase for the Sit-to-stand, Forward, and Stand-to-sit phases.

When trained to discriminate PD-MCI vs PD-NC, the algorithm presented a percent accuracy ranging from 80 to 87 percent during the training phase, proving to be very efficient in the inference phase on the test set for the Sit-to-stand, Mid-turning, and Turning-before-sitting phases.

When trained to discriminate Fluctuating vs Stable response to therapy, although it reached accuracies above 70 percent during validation across all phases, its accuracy dropped during testing for all phases except for Turning-before-sitting (72 percent of accuracy).

Finally, we tested the ANN classification algorithm. When trained to discriminate PD vs Controls, although the algorithm presented a percent accuracy of 91 during the training phase, in the inference phase it did not manage to discriminate classes correctly. However, in the following selected phases, it was the best performing model, achieving test accuracies of 85 percent in Mid-turning and 97 percent in Backward.

When trained to discriminate PD-MCI vs PD-NC, the algorithm failed to discriminate classes correctly in the inference phase although presented a percent accuracy ranging from 93 to 95 percent during the training phase. Despite this, the ANN performed well in most phases, achieving high test accuracies of 71 percent in Forward, 88 percent in Backward, and 69 percent in Stand-to-sit.

When trained to discriminate Fluctuating vs Stable response to therapy, it presented a percent accuracy ranging from 91 to 96 percent during the training phase; in the inference phase, it managed to correctly discriminate the classes. This suggests that the ANN is particularly effective in distinguishing different responses to therapy. It outperformed other models in the Mid-turning, Backward, and Stand-to-sit phases, achieving test accuracies of 83, 95, and 94 percent, respectively.

Sub-analysis results based on specific phases of iTUG are displayed in Table 2, showing high accuracy of trained models during the inference phase.

3.3. Machine learning algorithms performances based on EEG data: Training and validation

The KNN classification algorithm, when trained to discriminate PD vs Controls, exhibited a percent accuracy of 85 during the training phase; in the inference phase it succeeded to discriminate the classes correctly.

When trained to discriminate PD-MCI vs PD-NC, the algorithm exhibited a percent accuracy of 83 during the training phase; in the inference phase it did not manages to discriminate classes correctly.

When trained to discriminate Fluctuating vs Stable response to therapy, the algorithm exhibited a percent accuracy of 57 during the training phase; in the inference phase it did not manages to discriminate classes correctly.

The RFC algorithm, when trained to discriminate PD vs Controls, exhibited a percent accuracy of 76 during the training phase; in the inference phase, it did not manage to discriminate classes correctly.

When trained to discriminate PD-MCI vs PD-NC, the algorithm exhibited a percent accuracy of 96 during the training phase; it successfully and correctly discriminated the classes.

When trained to discriminate Fluctuating vs Stable response to therapy, the algorithm exhibited a percent accuracy of 91 during the training phase; in the inference phase, it successfully and correctly

Table 2
Machine learning models on iTUG data: performances and accuracies based on iTUG phases.

iTUG PHASEs	MODEL	ACCURACY (percent)
iTUG PD vs CTRL (A)		
Sit-to-stand	RFC	89
Forward	RFC	88
Mid-turning	ANN	85
Backward	ANN	97
Turning-before-sitting	KNN	81
Stand-to-sit	RFC	93
iTUG PD-MCI vs PD-NC (B)		
Sit-to-stand	RFC	75
Forward	ANN	71
Mid-turning	RFC	65
Backward	ANN	88
Turning-before-sitting	RFC	78
Stand-to-sit	ANN	69
iTUG PD Fluctuating vs Stable (C)		
Sit-to-stand	KNN	88
Forward	KNN	91
Mid-turning	ANN	83
Backward	ANN	95
Turning-before-sitting	RFC	72
Stand-to-sit	ANN	94

Legend: PD-NC (PD with Normal Cognition); PD-MCI (PD with Mild Cognitive Impairment); CTRL (Controls). Reference: historical data from healthy controls group-matched by age for both iTUG and EEG study (for iTUG: N = 31 subjects; for EEG: N = 27 subjects).

discriminated the classes.

The ANN classification algorithm, when trained to discriminate PD vs Controls, exhibited a percent accuracy of 85 during the training phase; in the inference phase, it did not manage to discriminate classes correctly.

When trained to discriminate PD-MCI vs PD-NC, the algorithm exhibited a percent accuracy of 75 during the training phase; in the inference phase, it did not manage to discriminate classes correctly.

When trained to discriminate Fluctuating vs Stable response to therapy, the algorithm exhibited a percent accuracy of 86 during the training phase; in the inference phase, it did not manage to discriminate classes correctly.

Data about best models inferences based on EEG raw data are displayed in Table 3.

3.4. Software development using models to analyze raw data

Through a process of thorough evaluation and careful analysis of the results obtained, we were able to identify the most suitable models for each phase of the iTUG and type of EEG analysis.

A Python software was developed to incorporate all the trained models for each analyzed phase. Specifically, the total number of models integrated into the software was 21, each specifically tailored for a particular dataset outcome. The software enabled the processing of raw data located within specified folders, pre-processed the data, subdivided the dataset to prepare it for inference, and executed the operations accordingly. The software's graphical user interface allowed users to select whether to run only iTUG, only EEG, or both, depending on the available data for diagnosis. Additionally, once initiated, the software displayed two types of results:

Pointwise results: these show the positive recurrence for each type of analysis conducted. Recurrence is understood as the percentage of the analyzed dataset that positively responds to the inference, or in other terms, the percentage of positive or negative bias beyond the midpoint, representing non-discrimination.

Overall results: another comprehensive result is shown as an average, considering all pointwise biases. This is done for both iTUG and EEG data, providing direct feedback on the diagnosed outcome.

The software leveraged various Python libraries and frameworks for its implementation. Specifically, it utilized libraries such as *scikit-learn* for machine learning model integration, pandas for data manipulation, and *matplotlib* or *seaborn* for visualization purposes. The GUI components were implemented using libraries like *tkinter* or *PyQt*, which facilitated the development of graphical interfaces in Python.

4. Discussion

4.1. Best model selection for iTUG and EEG raw data

Once the data collection and preparation phase regarding PD patients was completed, we faced the challenge of selecting the most suitable models for analyzing different aspects of the disease. Through standardized methodical approach, we examined a range of machine learning models to identify those with the best performance for each phase of the iTUG and EEG analyses.

Study first outcome was the accuracy of the selected models.

Table 3
Machine learning models on EEG data: performances and accuracies.

EEG	MODEL	ACCURACY (percent)
PD vs CTRL	KNN	98
PD-MCI vs PD-NC	RFC	88
PD Fluctuating vs Stable	RFC	89

Legend: PD-NC (PD with Normal Cognition); PD-MCI (PD with Mild Cognitive Impairment); CTRL (Controls).

Considering overall iTUG performances, algorithms demonstrated to correctly perform in discriminating the different classes we selected, with higher obtained accuracy for classes identifying PD motor response on validation phase.

Upon reviewing the results obtained from our analysis, we also noted that some models demonstrated higher performance in specific phases of the iTUG. For instance, in the backward phases, ANN Keras-based models achieved the highest accuracies in discriminated all classes (PD vs CTR, PD-MCI vs PD-NC and PD "fluctuating" vs "stable" based on motor response to L-dopa), indicating that these machine learning techniques, extensively used on literature for motion transducers data, may be particularly effective in predicting individual patient status.

Regarding the EEG analysis, results showed that KNN models achieved highest accuracy in discriminating PD vs controls, while RFC models achieved highest accuracy in discriminate PD patients based on cognitive impairment and response to treatment, suggesting that complexity of EEG data may require a more sophisticated model to obtain accurate results.

4.2. Limits of the applied machine learning methods

To avoid overfitting and ensure the models' ability to generalize effectively, it is essential to conduct rigorous external validation, which will help assess their true clinical accuracy and applicability. Additionally, we recognize several challenges concerning machine learning algorithms applications, such as the careful tuning of hyperparameters, the risk of distributional mismatch between training and test data subsets, and the variability of results when applied to non-stationary data. Despite these challenges, the models demonstrated valuable insights into the underlying patterns of the data, suggesting that with further refinement and validation, machine learning approaches could significantly enhance clinical decision-making [25–27].

In this study, we focused on binary classification to clearly differentiate between classes. This approach allowed us to simplify the analysis and focus on the primary clinical question. In future work, we plan to explore multiclass classification to capture more nuanced distinctions, such as between different stages of PD or other conditions, which would enhance the generalizability and applicability of the models.

Finally, the limited number of inner layers adopted in neural networks architectures may lead to a low capacity in representing analyzed data.

5. Conclusion

In the present study we evaluated machine algorithms applied to multimodal data, specifically motion and EEG signals, in order to differentiate PD from controls, but also to identify patients who can acutely respond to treatment or present initial cognitive decline. ANN-based models revealed the best performances when applied to specific phases of the iTUG. KNN and RFC-based model revealed best performances when applied to EEG data.

We then developed a software to test algorithms in independent samples for future validation tests. As a first step, we have developed a Python-based GUI to facilitate the use of our models and streamline data analysis. Looking ahead, we plan to expose APIs to enhance accessibility and integration with other tools and workflows. This future development will ensure broader usability and practical application of our findings. Future research may employ also deep learning architectures and Transformers [28,29], even method actually require a large amount of data to be analyzed. We aim to implement and optimize presented methods in future studies.

Ethics approval and consent to participate

the project is part of the IRMA study project, which has been approved by the Local Ethic Committee (Policlinico, Catania 1).

Author Agreement Statement

We the undersigned declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We understand that the Corresponding Author is the sole contact for the Editorial process.

He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs

Consent for publication

Written informed consent was obtained from study participants.

Funding

the project has been supported by funding related to the project: ”IRMA - Parkinson Cyclone in Life” - PO FESR 2014/2020.

CRediT authorship contribution statement

Zappia Mario: Writing – review & editing, Supervision, Project administration, Methodology, Data curation, Conceptualization. **Nicoletti Alessandra:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Donzuso Giulia:** Writing – review & editing, Supervision. **Mostile Giovanni:** Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Contrafatto Federico:** Writing – review & editing, Supervision, Resources, Data curation. **Quattropani Salvatore:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation. **Cicero Calogero Edoardo:** Writing – review & editing, Supervision. **Chiara Alessandra:** Writing – review & editing, Data curation. **Caci Michelangelo Riccardo:** Writing – review & editing, Data curation. **Terravecchia Claudio:** Writing – review & editing, Supervision, Data curation.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The project has been supported by funding related to the project: ”IRMA - Parkinson Cyclone in Life” - PO FESR 2014/2020. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.02.019.

References

[1] Tolosa E, Garrido A, Scholz SW, Poewe W. Challenges in the diagnosis of Parkinson’s disease. Lancet Neurol 2021;20:385–97.
[2] Mirelman A, Bonato P, Camicioli R, Ellis TD, Giladi N, Hamilton JL, Hass CJ, Hausdorff JM, Pelosin E, Almeida QJ. Gait impairments in Parkinson’s disease. Lancet Neurol 2019;18:697–708.
[3] Aarsland D, Batzu L, Halliday GM, Geurtsen GJ, Ballard C, Ray Chaudhuri K, Weintraub D. Parkinson disease-associated cognitive impairment. Nat Rev Dis Prim 2021;7:47.
[4] Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, Obeso J, Marek K, Litvan I, Lang AE, Halliday G, Goetz CG, Gasser T, Dubois B, Chan P, Bloem BR, Adler CH, Deuschl G. MDS clinical diagnostic criteria for Parkinson’s disease. Mov Disord 2015;30:1591–601.
[5] Nonnekes J, Timmer MH, de Vries NM, Rascol O, Helmich RC, Bloem BR. Unmasking levodopa resistance in Parkinson’s disease. Mov Disord 2016;31:1602–9.
[6] Vaillancourt DE, Schonfeld D, Kwak Y, Bohnen NI, Seidler R. Dopamine overdose hypothesis: evidence and clinical implications. Mov Disord 2013;14:1920–9.
[7] Di Lazzaro G, Ricci M, Al-Wardat M, Schirinzi T, Scalise S, Giannini F, Mercuri NB, Saggio G, Pisani A. Technology-Based Objective Measures Detect Subclinical Axial Signs in Untreated, de novo Parkinson’s Disease. J Park Dis 2020;10:113–22.
[8] Mostile G, Contrafatto F, Terranova R, Terravecchia C, Luca A, Sinitò M, Donzuso G, Cicero CE, Sciacca G, Nicoletti A, Zappia M. Turning and Sitting in Early Parkinsonism: Differences Between Idiopathic Normal Pressure Hydrocephalus Associated with Parkinsonism and Parkinson’s Disease. Mov Disord Clin Pr 2022;10:466–71.
[9] Suppa A, Kita A, Leodori G, Zampogna A, Nicolini E, Lorenzi P, Rao R, Irrera F. l-DOPA and Freezing of Gait in Parkinson’s Disease: Objective Assessment through a Wearable Wireless System. Front Neurol 2017;8:406.
[10] Mostile G, Giuliano L, Monastero R, Luca A, Cicero CE, Donzuso G, Dibilio V, Baschi R, Terranova R, Restivo V, Sofia V, Zappia M, Nicoletti A. Electrocortical networks in Parkinson’s disease patients with Mild Cognitive Impairment. The PaCoS study. Park Relat Disord 2019;64:156–62.
[11] Gérard M, Bayot M, Derambure P, Dujardin K, Defebvre L, Betrouni N, Delval A. EEG-based functional connectivity and executive control in patients with Parkinson’s disease and freezing of gait. Clin Neurophysiol 2022;137:207–15.
[12] Di Lazzaro G, Ricci M, Al-Wardat M, Schirinzi T, Scalise S, Giannini F, Mercuri NB, Saggio G, Pisani A. Technology-Based Objective Measures Detect Sub-clinical Axial Signs in Untreated, de novo Parkinson’s Disease. J Park Dis 2020;10:113–22.
[13] Castelli Gattinara Di Zubiena F, Menna G, Mileti I, Zampogna A, Asci F, Paoloni M, Suppa A, Del Prete Z, Palermo E. Machine Learning and Wearable Sen- sors for the Early Detection of Balance Disorders in Parkinson’s Disease. Sens (Basel) 2022;22:9903.
[14] Trabassi D, Serrao M, Varrecchia T, Ranavolo A, Coppola G, De Icco R, Tassorelli C, Castiglia SF. Machine Learning Approach to Support the Detection of Parkinson’s Disease in IMU-Based Gait Analysis. Sens (Basel) 2022;22:3700.
[15] Sigcha L, Borz` L, Amato F, Rechichi I, Ramos-Romero C, Cárdenas A, Gasco L, Olmo G. Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson’s disease: A systematic review. Expert Syst Appl 2023;229(Part A):120541.
[16] Dibilio V, Nicoletti A, Mostile G, Toscano S, Luca A, Raciti L, Sciacca G, Vasta R, Cicero CE, Contrafatto D, Zappia M. Dopaminergic and non-dopaminergic gait components assessed by instrumented timed up and go test in Parkinson’s disease. J Neural Transm (Vienna) 2017;124:1539–46.
[17] Mostile G, Terranova R, Rascuna C, Terravecchia C, Cicero CE, Giuliano L, Dav`ı M, Chisari C, Luca A, Preux PM, Jankovic J, Zappia M, Nicoletti A. Clinical-Instrumental patterns of neurodegeneration in Essential Tremor: A data-driven approach. Park Relat Disord 2021;87:124–9.
[18] Mostile G, Nicoletti A, Dibilio V, Luca A, Pappalardo I, Giuliano L, Cicero CE, Sciacca G, Raciti L, Contrafatto D, Bruno E, Sofia V, Zappia M. Electroencephalographic lateralization, clinical correlates and pharmacological response in untreated Parkinson’s disease. Park Relat Disord 2015;21:948–53.
[19] Mostile G, Giuliano L, Dibilio V, Luca A, Cicero CE, Sofia V, Nicoletti A, Zappia M. Complexity of electrocortical activity as potential biomarker in untreated Parkinson’s disease. J Neural Transm (Vienna) 2019;126:167–72.
[20] Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE. Systematic review of levodopa dose equivalency reporting in Parkinson’s disease. Mov Disord 2010;25:2649–53.
[21] Litvan I, Goldman JG, Tröster AI, Schmand BA, Weintraub D, Petersen RC, Mollenhauer B, Adler CH, Marder K, Williams-Gray CH, Aarsland D, Kulisevsky J, Rodriguez-Oroz MC, Burn DJ, Barker RA, Emre M. Diagnostic criteria for mild cognitive impairment in Parkinson’s disease: movement Disorder Society Task Force guidelines. Mov Disord 2012;27:349–56.
[22] Linkenkaer-Hansen K, Nikouline VV, Palva JM, Ilmoniemi RJ. Long-range temporal correlations and scaling behavior in human brain oscillations. J Neurosci 2001;21:1370–7.
[23] Schirrmeister RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W, Ball T. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum Brain Mapp 2017;38:5391–420.
[24] Sabbagh D, Ablin P, Varoquaux G, Gramfort A, Engemann DA. Predictive regression modeling with MEG/EEG: from source power to signals and cognitive states. Neuroimage 2020;222:116893.
[25] Kubota KJ, Chen JA, Little MA. Machine learning for large-scale wearable sensor data in Parkinson’s disease: Concepts, promises, pitfalls, and futures. Mov Disord 2016;31:1314–26.
[26] Ramdhani RA, Khojandi A, Shylo O, Kopell BH. Optimizing clinical assess- ments in Parkinson’s disease through the use of wearable sensors and data driven modeling. Front Comput Neurosci 2018;12:72.
[27] Chandrabhatla AS, Pomeraniec IJ, Ksendzovsky A. Co-evolution of machine learning and digital technologies to improve monitoring of Parkinson’s disease motor symptoms. NPJ Digit Med 2022;5:32.
[28] Shavit Y, Klein I. Boosting inertial-based human activity recognition with transformers. IEEE Access 2021;9:53540–7.
[29] Ahmadian S, Rostami M, Farrahi V, Oussalah M. A novel physical activity recognition approach using deep ensemble optimized transformers and reinforcement learning. Neural Netw 2024;173:106159.