Guimarães *et al. BMC Medical Informatics and Decision Making* (2025) 25:157
https://doi.org/10.1186/s12911-025-02979-9

BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Detecting cognitive impairment in cerebrovascular disease using gait, dual tasks, and machine learning

Vânia Guimarães[1,2*], Inês Sousa[1] and Miguel Velhote Correia[2,3]

Abstract

Background Cognitive impairment is common after a stroke, but it can often go undetected. In this study, we investigated whether using gait and dual tasks could help detect cognitive impairment after stroke.

Methods We analyzed gait and neuropsychological data from 47 participants who were part of the Ontario Neurodegenerative Disease Research Initiative. Based on neuropsychological criteria, participants were categorized as impaired ($n=29$) or cognitively normal ($n=18$). Nested cross-validation was used for model training, hyperparameter tuning, and evaluation. Grid search with cross-validation was used to optimize the hyperparameters of a set of feature selectors and classifiers. Different gait tests were assessed separately.

Results The best classification performance was achieved using a comprehensive set of gait metrics, measured by the electronic walkway, that included dual-task costs while performing subtractions by ones. Using a Support Vector Machine (SVM), we could achieve a sensitivity of 96.6%, and a specificity of 61.1%. An optimized threshold of 27 in the Montreal Cognitive Assessment (MoCA) revealed lower classification performance than the gait metrics, although differences in classification results were not significant. Combining the classifications provided by MoCA with those provided by gait metrics in a majority voting approach resulted in a higher specificity of 72.2%, and a high sensitivity of 93.1%.

Conclusions Our results suggest that gait analysis can be a useful tool for detecting cognitive impairment in patients with cerebrovascular disease, serving as a suitable alternative or complement to MoCA in the screening for cognitive impairment.

Keywords Cerebrovascular disease, Cognitive impairment, Dual-task, Gait, Machine learning, Stroke

*Correspondence:
Vânia Guimarães
vania.guimaraes@aicos.fraunhofer.pt
[1]Fraunhofer Portugal AICOS, Rua Alfredo Allen 455/461, Porto 4200-135, Portugal
[2]Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, Porto 4200-465, Portugal
[3]INESC TEC, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, Porto 4200-465, Portugal

Background

Cognitive impairment (CI) due to cerebrovascular disease (CVD), also known as vascular cognitive impairment (VCI) is common after stroke. It affects mostly attention and executive functions [1, 2], ranging from mild to major CI or dementia [1, 3, 4]. Vascular dementia, which is the second most common cause of dementia after Alzheimer's disease [1, 5, 6], can significantly impact patients' quality of life and independence [5, 7], being associated with increased mortality, disability, and institutionalization [8].

Despite significant progress in stroke care in recent decades, CI after stroke remains a common and underdiagnosed problem [7, 9, 10]. Studies estimate that 30–50% of stroke survivors experience CI [7], but recent systematic reviews suggest that the prevalence may be even higher, with more than half of stroke survivors experiencing some degree of CI [11]. In the first year after a stroke, the prevalence of mild CI ranges from 17.5% to 54.9%, with a pooled prevalence of 38% [12]. As the population ages and the number of stroke survivors increases, effective management of post-stroke CI will become increasingly important for both patients and caregivers [9].

The first step in the management of CI is diagnosis. The gold standard for diagnosis is a comprehensive neuropsychological examination [9] that is not feasible at a population level. In practice, a two-step approach is advised, starting with a primary screening test such as the Montreal Cognitive Assessment (MoCA) [4, 8]. While MoCA is commonly used for this purpose, its low specificity, and moderate to good sensitivity for the diagnosis of post-stroke CI coupled with a lack of standardization still limit its use [6, 9].

Researchers have been exploring alternative methods to improve the screening of CI. Gait analysis, for instance, was shown to distinguish normal aging from mild CI in non-stroke patients [13]. Further research showed that gait performance is affected by cognitive load in people with CI due to cognitive-motor, or, dual-task, interference [14, 15]. A recent meta-analysis showed that dual-task gait characteristics allow better differentiation between groups, with an increased sensitivity for the detection of mild CI [13].

Because attention is often impaired, patients with VCI may struggle with multitasking [9]. Studies suggest that performance on dual tasks can be compromised in patients with VCI [16], and suggest that gait assessment under dual-task conditions could be a useful tool for early detection of VCI [17]. Dual-task assessment is attractive because it is less influenced by educational level, is practical, fast, and easy to administer in clinical practice [18], and may be a good alternative to screening tests like MoCA.

Digital biomarker technologies, often combined with machine learning-based predictive models, are increasingly explored for early detection of CI in community settings [19]. Support Vector Machines (SVMs) are often used for this purpose by analyzing gait characteristics [19]. For example, Boettcher et al. [20] used dual-task gait data collected from a computerized walkway to differentiate subjects with mild CI from healthy individuals, achieving an accuracy of 77.2%. Aoki et al. [18] used the Microsoft Kinect sensor to capture dual-task gait, achieving an AUC of 74.7% in differentiating subjects with different levels of cognitive performance, as measured by the Mini-Mental State Examination (MMSE). Ghoraani et al. [21] employed an SVM in a one-vs-one manner with a majority vote to differentiate healthy subjects, older adults with mild CI, and older adults with Alzheimer's disease. An accuracy of 86.0% was achieved in the differentiation of cognitively impaired and healthy subjects using gait features extracted from an electronic walkway. Shahzad et al. [22] used an inertial sensor-based gait analysis and a machine learning framework to distinguish mild CI due to Alzheimer's disease from healthy subjects, finding that dual-task walking provided a better distinction than single-task, with a classification accuracy of 70.0%. Finally, Jung et al. [23] used sequential gait characteristics extracted from inertial sensors coupled with a long short-term memory network to categorize community-dwelling older adults into three groups based on their scores of the MMSE; they achieved improved F1 scores of 97.4% using temporal gait features extracted from usual and fast pace walking trials. Taken together, these studies suggest that machine learning and gait assessments have the potential to be objective tools for cognitive screening that do not heavily rely on cognitive testing.

Although some research has investigated the use of gait characteristics to detect CI, it is unclear whether machine learning and gait assessments can effectively distinguish CI from cognitively normal in post-stroke patients. In this study, we aim to objectively evaluate the clinical utility of gait in classifying CI in post-stroke patients using a machine learning-based approach. We used as input data from different gait tests, including dual tasks and fast walking conditions. We also tested multiple machine learning classifiers and two different gait assessment technologies that provided varying levels of detail about gait metrics. Finally, we objectively compared the classification results with those obtained using MoCA and proposed a majority voting approach to combine MoCA with gait. We hypothesized that: i) machine learning models based on gait metrics differentiate VCI from healthy individuals with a performance comparable to MoCA; ii) the dual-tasks provide better classification results than the single-tasks; iii) a more comprehensive analysis of gait allows better differentiation of the groups than simpler gait metrics; and iv) the combination of MoCA with gait metrics leads to better performances than using each modality alone.

Materials and methods

Dataset

The data used in this study were obtained from the Ontario Neurodegenerative Disease Research Initiative (ONDRI) [24, 25]. ONDRI is a multi-site prospective cohort study investigating Alzheimer's dementia and mild CI (AD and MCI), amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTD), Parkinson's disease (PD), and cerebrovascular disease (CVD). The ONDRI protocol includes assessments for genomics, neuroimaging, ocular function, gait and balance, and neuropsychological testing. In the present study, we have only used the baseline data from the cerebrovascular disease (CVD) cohort, including demographic, clinical, neuropsychological, and gait data.

Participants in the CVD group experienced a mild to moderate ischemic (or silent) stroke event verified on neuroimaging 3 or more months before enrollment. They also met the following inclusion criteria: (a) age between 55 and 85 years old, (b) proficient in speaking and understanding English, (c) 8 or more years of formal education, (d) mild-moderate stroke severity defined by scores 0–3 in the modified Rankin scale, (e) a MoCA score of at least 18. Exclusion criteria included vascular cause of symptoms, large cortical strokes, severe CI, aphasia, history of dementia prior to the stroke, inability to write, and/or severe functional disability limiting ability to perform assessments [24, 26, 27].

The current study received approval from the Ethics Committee of the Faculty of Medicine of the University of Porto (51/CEFMUP/2022). Access to the ONDRI Foundational Study Data was granted after approval by Brain-CODE's Data Access Committee [28].

Assessments

Demographic and clinical data

The dataset included demographic data, such as age, education (reported in years), height, weight, and leg length. Additionally, it included a rate of depression, as assessed by the Quick Inventory of Depressive Symptomatology (QIDS) [29].

Study partners rated the participant's ability to function independently across eight instrumental activities of daily living (iADLs; i.e., telephone use, shopping, food preparation, housekeeping, laundering, use of transportation, medication management, and financial management) and six basic ADLs (i.e., feeding, dressing, grooming, ambulating, bathing, and toileting). The percent of independence on relevant items was provided for iADLs and ADLs.

MoCA

MoCA is a paper-and-pencil screening tool for CI [30]. It evaluates multiple cognitive domains, including, executive function, memory, language, visuospatial ability, orientation, attention, concentration, and working memory. The score of MoCA is corrected for low education (≤12 years) by adding an extra point, having a maximum score of 30 points [30]. With a high sensitivity of 83–97%, MoCA is currently the preferred tool for the screening of CI at primary care level [31]. The cutoff of <26 is commonly used to detect CI [8, 27, 30].

Neuropsychological assessment

The neuropsychological assessment consisted of a standardized battery administered to all participants in the ONDRI study. The ONDRI study followed standard quality assurance and quality control procedures, as described in [32], ensuring the rigor and accuracy of the data.

To characterize areas of CI in individual participants, the tests were categorized into five cognitive domains, as proposed by [33] after a consensus agreement among the ONDRI Clinical Neuropsychologists. In addition to raw scores from the neuropsychological tests, the ONDRI dataset provided standard scores (z-scores, t-scores, percents, or scaled scores) based on published normative data (education- and/or age-adjusted). Normative values were not available in two of the considered tests (BDAE: Semantic probe and BVMT-R: Copy trial) and they were, thus, excluded from the analysis. The resulting cognitive domains and associated test measures are shown in Table 1.

We categorized participants as impaired or as cognitively normal using the comprehensive criteria proposed by [34]. According to this criteria, individuals were classified as cognitively normal if, at most, performance on one measure within one or two cognitive domains fell more than 1 SD below age/education-appropriate norms [34].

Before applying the comprehensive criteria, we filled in missing values using k-Nearest Neighbors (KNN). Missing neuropsychology data were imputed using the mean value from the 5 nearest neighbors weighted by the inverse of their Euclidean distance. Since KNN is based on distances, before imputation, we standardized the data, i.e. removed their mean and scaled to unit variance, to generate unbiased estimations of missing values [35]. After imputation, the values were transformed to the original scale.

Table 1 Neuropsychological assessment

Cognitive domains	Tests included
Attention and working memory	Symbol digit modality test (coding)
	Trail making test–Part A (time)
	WAIS-III: Digit span forward (longest span)
	WAIS-III: Digit span backward (longest span)
	WAIS-III: Digit span total
	DKEFS: Color naming (time)
	DKEFS: Word reading (time)
Executive function	Trail making test–Part B (time)
	DKEFS: Interference (time)
	DKEFS: Inhibition/switching (time)
	DKEFS: Letter fluency
	DKEFS: Category fluency
	WASI-II: Matrix
Language	Boston naming – 15 item (pro-rated)
	TAWF: Verb naming
	BDAE: Semantic probe (excluded)
	WASI-II: Vocabulary
Verbal memory	RAVLT: Immediate
	RAVLT: Long-delay
	RAVLT: Recognition discrimination
Visuospatial awareness	Judgement of line orientation
	VOSP: Incomplete letters
	BVMT-R: Copy trial (raw) (excluded)

BDAE: Boston Diagnostic Aphasia Examination; BVMT-R: Brief Visuospatial Memory Test-Revised; DKEFS: Delis-Kaplan Executive Function System; RAVLT: Rey Auditory Verbal Learning Test; TAWF: Test of Adolescent/Adult Word Finding; VOSP: Visual Object and Space Perception Battery; WAIS-III: Wechsler Adult Intelligence Scale–Third Edition; WASI-II: Wechsler Abbreviated Scale Intelligence–Second Edition

Gait

Quantitative gait parameters were assessed using wearable inertial sensors (Gulf Coast Inc.; Shimmer Inc.) worn bilaterally on the ankles and at the hip. In two study sites, gait performance was additionally assessed using electronic walkway systems (GAITRite® or PKMas®) [24, 36]. To facilitate a comparison between inertial sensors and electronic walkway systems, we have only considered the participants who were assessed by the two systems.

Gait was assessed under three conditions, using the following order: (1) preferred walking speed, or single-task walking (SS), (2) dual-task walking (DT), and (3) fast walking (F). The secondary tasks were administered in the following order: (1) subtracting serial ones from 100 (DS1), (2) naming animals (DAn), and (3) subtracting serial sevens from 100 (DS7). The cognitive tasks were performed out loud, without any prioritization instruction. All participants walked along a 6-m path, starting and finishing 1 m away to ensure that only steady-state walking was captured [24, 36].

The preferred walking speed condition was repeated three times, and average values were calculated. Gait characteristics such as speed, cadence, and total number of steps were reported for the participants assessed with the accelerometer. The participants assessed using the electronic walkways had additional gait characteristics, including, gait speed, stride velocity, step and stride time, step and stride length, double support time, swing time, step and stride width, cadence, and corresponding variabilities, calculated using the coefficient of variation, CoV, obtained from [(standard deviation/mean) ×100] [36].

Since some of the metrics extracted from the electronic walkway provide similar information (e.g., stride time and cadence, step and stride metrics), we excluded step metrics and cadence from the analysis. The metrics included were, thus, stride length, double support time, stride time, stride velocity, stride width, swing time, and their respective variabilities.

For all these metrics we calculated dual-task costs (DTC, in %) using the formula [(single-task metric—dual-task metric)/single-task metric] ×100, which quantifies the magnitude of the effect of cognitive load on gait performance [36]. We also calculated the capacity index comparing preferred speed with fast walking trials using the formula [(fast metric—single-task metric)/single-task metric] ×100.

Machine learning pipeline

We developed a machine learning pipeline for classifying participants into normal or cognitively impaired using the metrics from gait tests. We assessed each gait test separately (i.e., SS, DAn, DS1, DS7, and F) to compare each test's ability to differentiate both groups. Additionally, we tested the combinations of dual-task tests (and fast-walking test) with SS, i.e., DAn + SS, DS1 + SS, DS7 + SS, and F + SS. To this purpose, we analyzed and reported the best-performing experiment from the following three possibilities: (a) using dual-task costs (or capacity indexes) alone, (b) combining dual-task costs (or capacity indexes) and SS metrics, and (c) combining dual-task costs (or capacity indexes) and dual-task (or fast-walking) test metrics. The five gait tests and the four combinations resulted in a total of nine reported experiments.

The experiments were performed using metrics provided by the accelerometer and repeated using the metrics provided by the electronic walkway. Since our goal was to compare the limited set of features provided in this study by the accelerometer with the more comprehensive set provided by the electronic walkway (described in the previous section), we have only included participants assessed by the two systems.

One participant in the cognitively normal group had no accelerometer data on DAn, DS7 and F tests and was, as such, excluded from experiments including these data. The same happened with one participant in the cognitively impaired group that had no accelerometer data on the DAn test. All the remaining variables were present, yielding 29 participants in the impaired group (or 28 when data were missing) and 18 in the cognitively normal group (or 17 when data were missing).

To avoid overly optimistic estimations and account for the limited number of participants in this study, we used a nested cross-validation (CV) approach [37], which consists of an outer and inner CV loop, as shown in Fig. 1. Although leave-one-subject-out (LOSO) cross-validation is a common approach for small datasets, we opted for a k-fold split with $k = 20$ in nested CV to reduce computational load while maintaining sufficient training data and robust model evaluation. The dataset was, thus, first divided into 20 equally-sized outer folds, yielding 42–45 participants' data for hyperparameter tuning and model training (training sets) and 2–3 participants' data for evaluation (test sets). In the inner loop, the outer training data were further divided into 5 folds, forming the inner training and validation folds. The inner loops were used for hyperparameter tuning in a grid search. When the best set of hyperparameters was found, the whole (outer) training set was used to retrain the model. The model was then used to classify the test set in the outer loop.

As illustrated in Fig. 1, the machine learning pipeline consists of: (1) preprocessing, (2) feature selection, (3) sampling, and (4) classification. The preprocessing step consisted of a z-score standardization that scaled the features to zero mean and unit standard deviation. For the feature selection, we tested different methods: we optimized the number of features to select (1, 2, 3, 5, or 8) using the highest-scored features (KBest) according to their mutual information (or mutual dependence) with the target [38], or using recursive feature elimination (RFE) based on the feature ranking provided by a Support Vector Machine (SVM) with a linear kernel; we also tested feature selection based on feature importance as provided by Random Forest (RF) [39] with 50 estimators, selecting all those features with importance above the mean. We also tested the pipeline without any feature selection. In the sampling step, we used the Synthetic Minority Oversampling Technique (SMOTE) algorithm [40]. Since the incidence of cognitively normal participants in the dataset was only 38.3%, SMOTE was used to create new synthetic examples of the minority class. According to [40] SMOTE provides better results when applied after feature selection. We also evaluated the performance of the classifiers without applying any resampling strategy, as all the classifiers tested in the subsequent stage were able to adjust for dataset imbalance by using class weighting. In the classification stage, we optimized the following hyperparameters in three different classifiers: the regularization strength, C (0.001, 0.01, 0.1, 1.0, 10, or 100) and kernel (linear or rbf) of the Support Vector Machine (SVM); the number of estimators (10, 50, or 100) of the Random Forest (RF); and the regularization strength, C (0.001, 0.01, 0.1, 1.0, 10, or 100) of the Logistic Regression (LR) classifier.

The total number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were obtained for all test partitions in the outer loop and used to calculate the global metrics of performance, where the cognitively impaired group was considered the positive class. To evaluate and compare the performance of the machine learning models, we calculated balanced accuracy, sensitivity, specificity, F1 score, and the area under the receiver operating characteristic (ROC) curve (AUC). Since we were dealing with imbalanced data, balanced accuracy was used as the scoring criteria to select the

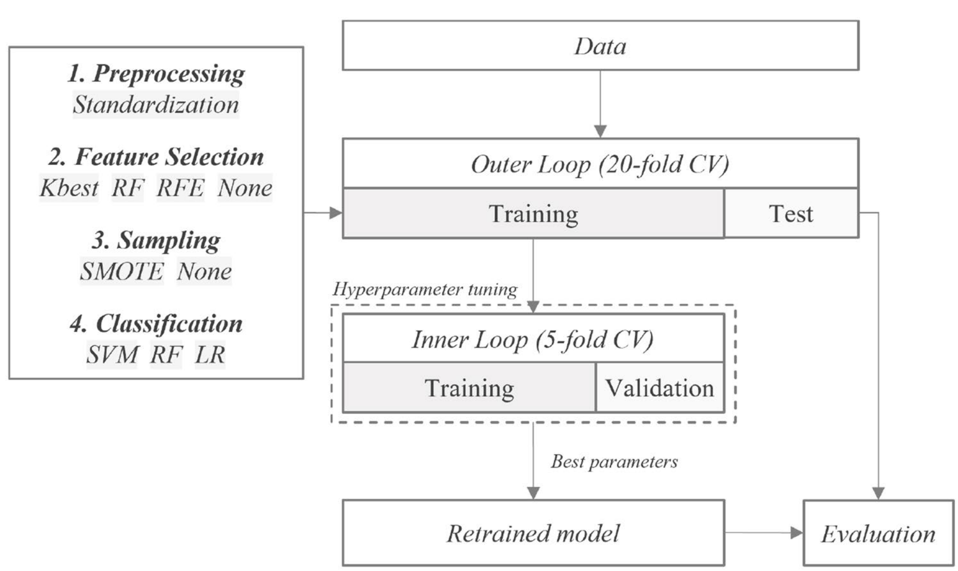

Fig. 1 Machine learning pipeline with nested cross-validation

best set of hyperparameters in the grid search. Balanced accuracy was also used to compare multiple machine-learning experiments.

Experiments were performed using Python 3 and the package scikit-learn v1.1.3 [41].

Classification using MoCA scores

We compared the performance of the machine learning experiments with MoCA. Two different experiments were conducted: (1) we used the classical threshold of 26 to identify the groups and (2) we used a ROC analysis to determine the optimal MoCA cut-point that maximized sensitivity and specificity in the current dataset. The optimal cut-point was considered the point of the ROC curve closest to the upper left corner (0,1), defined as that yielding the minimum value of $(1-sensitivity)^2 + (1-specificity)^2$ [42]. A third experiment combined the classifications obtained with MoCA with those obtained using gait metrics in two of the best-performing conditions. The classifications obtained by the three different approaches were combined using a majority voting scheme [43].

Table 2 Descriptive statistics ($n=47$)

Characteristic	Normal ($n=18$)	Impaired ($n=29$)	*p*-value
Age [years]	67.3 ± 5.1	69.6 ± 7.3	0.220
Height [cm]	174.7 ± 6.5	169.1 ± 10.7	**0.033***
Leg length [cm]	93.1 ± 6.2	91.8 ± 8.1	0.571
Weight [Kg]	85.7 ± 12.5	81.8 ± 14.2	0.242
Female, n (%)	4 (22.2)	8 (27.6)	0.098
Education [years]	15.3 ± 2.9	15.4 ± 2.7	0.781
Clinical tests			
QIDS score	4.0 ± 2.5	4.1 ± 3.0	0.833
iADLs [%]	95.9 ± 6.4	85.2 ± 18.1	**0.036***
ADLs [%]	99.5 ± 1.3	99.3 ± 2.5	0.949
MoCA score	27.6 ± 2.0	24.3 ± 2.8	**<0.001***
Gait using the accelerometer			
SS speed [m/s]	1.18 ± 0.28	0.98 ± 0.27	**0.024***
DAn speed [m/s]	1.05 ± 0.26	0.84 ± 0.26	**0.012***
DS1 speed [m/s]	1.05 ± 0.29	0.91 ± 0.28	0.112
DS7 speed [m/s]	0.96 ± 0.32	0.79 ± 0.25	0.106
F speed [m/s]	1.73 ± 0.49	1.45 ± 0.45	0.060
Gait using the electronic walkway			
SS speed [m/s]	1.25 ± 0.15	1.11 ± 0.20	**0.021***
DAn speed [m/s]	1.10 ± 0.24	0.92 ± 0.21	**0.007***
DS1 speed [m/s]	1.17 ± 0.21	1.02 ± 0.23	**0.024***
DS7 Speed [m/s]	1.15 ± 0.32	0.86 ± 0.19	**0.002***
F Speed [m/s]	1.67 ± 0.41	1.52 ± 0.27	0.088

Data are mean values ± standard deviation or the number of participants per category (absolute and relative frequency) when indicated. Group differences were evaluated using the Independent Samples *T*-test, Mann-Whitney U test, or Pearson's chi-square. *$p < 0.05$, *p*-values are two-tailed significance and bold values indicate significance. QIDS: Quick Inventory of Depressive Symptomatology; iADLs: Percent of independence in instrumental activities of daily living; ADLs: Percent of independence in activities of daily living; SS: Single-task; DAn: Walking while naming animals; DS1: Serial subtraction by 1s; DS7: Serial subtraction by 7s; F: Walking Fast

Statistical analysis

Demographic and medical characteristics were summarized using either means and standard deviations, or frequencies and percentages, as appropriate, to characterize both groups. The comparisons between groups were performed using the Independent Samples *T*-test or the Mann-Whitney U test for continuous variables, and Person's chi-square test (χ^2) for categorical variables. The Mann-Whitney U test was used when data were not normally distributed or when outliers were present outside ±1.5 of the interquartile range. To test for normality, we used the Shapiro-Wilk test.

To compare the results of the machine learning experiments, we used Paired Samples *T*-Tests or Wilcoxon signed-rank tests, that compared the average performance of the validation sets in all inner cross-validation (CV) runs. Wilcoxon signed-rank tests were used when paired data differences were not normally distributed or when outliers were present. Bonferroni-adjusted *p*-values were reported for the multiple comparisons [44].

McNemar's Test was used to compare the diagnosis provided by MoCA, and the classifications provided by machine learning experiments, as recommended in [45].

The statistical analysis was performed using Python 3 and the statistical tools from SciPy v1.9.3, NumPy v1.23.5, and statsmodels v0.13.5 [46]. A *p*-value of less than 0.05 indicated statistical significance.

Results

Descriptive statistics

Of the 161 participants with CVD available in the ONDRI dataset, only 47 had gait features measured by both the accelerometer and the electronic walkway. Of the 47 participants, 29 (61.7%) met the criteria for CI, whereas 18 (38.3%) were classified as cognitively normal. The descriptive statistics of the two groups are provided in Table 2.

Significant differences between groups were obtained on MoCA scores and iADLs percent, indicating lower global cognitive performance in the impaired group and less independence on iADLs. The difference in ADLs was not significant, denoting that both groups were still equally capable of performing these tasks. Additionally, there were no significant differences in depression scores, age, and education between the two groups.

The group with CI had a slower gait speed, as indicated by significant differences in multiple gait tests. The electronic walkway revealed significant differences in gait speed in more tests than the accelerometer. No significant differences between groups were obtained for gait speed while walking fast, as measured either by the accelerometer or the electronic walkway. Although the impaired group had significantly shorter height, leg length presented no significant differences between

Table 3 Performance on test sets using gait metrics from the accelerometer (in %)

Condition	SS	DAn	DS1	DS7	F	DAn + SS[a)]	DS1 + SS[b)]	DS7 + SS[c)]	F + SS[b)]
Classifier	LR	LR	SVM	LR	LR	RF	LR	LR	RF
Feature Sel.	None	RF	None	None	None	RFE	RFE	KBest	KBest
Sampling	None	SMOTE	SMOTE	None	None	None	None	SMOTE	None
# Features	3.0 ± 0.0	2.0±0.0	3.0±0.0	3.0±0.0	3.0±0.0	2.0±0.0	4.6±1.0	4.5±1.0	3.4±1.7
Bal. Acc.	69.5	64.5	62.6	61.2	60.4	66.9	**70.6**	68.1	66.1
Sensitivity	72.4	64.3	58.6	51.7	62.1	75.0	69.0	65.5	79.3
Specificity	66.7	64.7	66.7	70.6	58.8	58.8	72.2	70.6	52.9
F1 score	75.0	69.2	65.4	61.2	66.7	75.0	74.1	71.7	76.7
AUC	66.3	65.5	65.7	57.4	63.1	63.4	75.5	66.5	57.8

SS: Single-task; DAn: Walking while naming animals; DS1: Serial subtraction by 1s; DS7: Serial subtraction by 7s; F: Walking Fast; LR: Logistic Regression; RF: Random Forest; SVM: Support Vector Machine; RFE: Recursive feature elimination; AUC: Area under the curve. [a)] using dual-task costs; [b)] using dual-task costs (or capacity indexes) and SS metrics; [c)] using dual-task costs and dual-task metrics

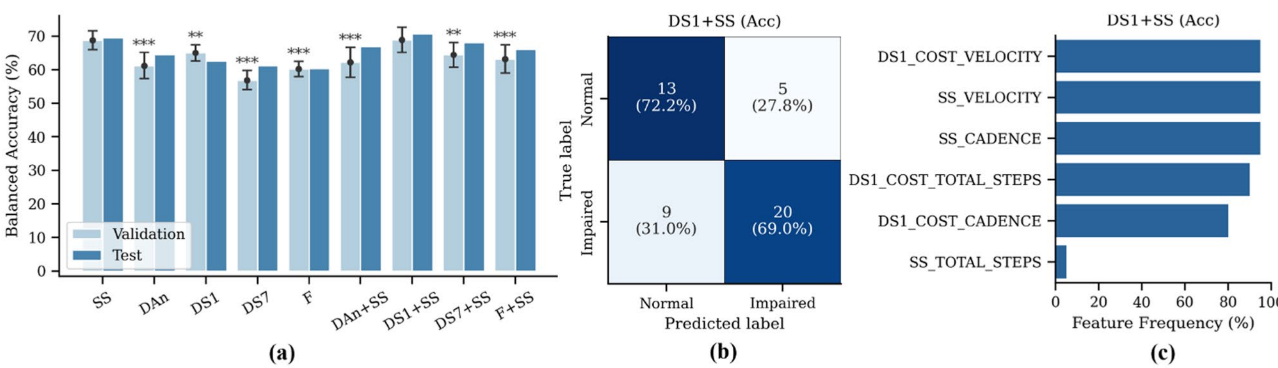

Fig. 2 Performance using features extracted from the accelerometer. (**a**) Performance on validation and test sets, where vertical lines indicate standard deviation of the 20 inner CVs. The statistical analysis compares the best-performing experiment (DS1 + SS) with the results achieved in the other conditions, using Paired Samples *T*-tests or Wilcoxon signed-rank tests with Bonferroni correction, where $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$ indicate statistical significance. (**b**) Confusion matrix of the DS1 + SS condition. (**c**) Feature selection frequency for the DS1 + SS condition

groups, which should not contribute to differences in gait performance [47].

Classification using accelerometer metrics

The best-performing classifiers, feature selectors, and sampling methods are shown in Table 3 for each tested gait condition using metrics from the accelerometer. The average number (and standard deviation) of features selected (within each outer training fold) are also shown in this table. Performance metrics are reported for the test set.

Balanced accuracy spanned from 60.4% (using F walking features) to 70.6% (using DS1 + SS walking features). The features extracted from the SS condition provided better performance than DAn, DS1, DS7, and F conditions, with 72.4% of sensitivity, 66.7% of specificity, and 66.3% of AUC. The performances of dual-task conditions (and F) improved when dual-task costs (or capacity indexes) were included (e.g., the performance of DAn + SS was higher than DAn alone). The best performing combination overall was with DS1 + SS, with a sensitivity of 69.0%, a specificity of 72.2%, and an AUC of 75.5% (Table 3).

The performance (balanced accuracy) on validation and test sets using gait metrics from the accelerometer is visually compared in Fig. 2a. Statistical results are provided in this figure comparing DS1 + SS with the other conditions.

From the analysis of Fig. 2a we verify that classification performance (measured on validation folds) of the best performing condition, DS1 + SS, although superior to the SS condition, does not differ significantly. All the remaining conditions had a lower average performance, differing significantly from the DS1 + SS condition.

We note that in the SS condition only 3.0 ± 0.0 features were included, whereas in the DS1 + SS condition an average of 4.6 ± 1.0 were included (Table 3). In the case of the accelerometer, as only three metrics were measured (i.e., number of steps, cadence, and speed), the maximum number of features selected was either three or six in the case of the experiments including dual-task costs. The features that were most commonly used in the DS1 + SS condition are shown in Fig. 2c. In this condition, the most relevant features were the cost in velocity, single-task velocity, single-task cadence, the cost in total steps, and the cost in cadence. The total number of steps in the single-task condition was the least relevant feature.

Table 4 Performance on test sets using gait metrics from the electronic walkway (in %)

Condition	SS	DAn	DS1	DS7	F	DAn + SS[a)]	DS1 + SS[a)]	DS7 + SS[b)]	F + SS[b)]
Classifier	LR	SVM	LR	SVM	RF	SVM	SVM	SVM	LR
Feature Sel.	RFE	RFE	RFE	None	RF	None	KBest	RF	RFE
Sampling	SMOTE	SMOTE	None	SMOTE	None	None	None	SMOTE	SMOTE
# Features	2.0 ± 2.2	3.2 ± 2.2	2.2 ± 1.8	12.0 ± 0.0	6.0 ± 0.8	12.0 ± 0.0	1.7 ± 2.1	8.6 ± 1.3	2.8 ± 2.1
Bal. Acc.	76.8	75.1	71.6	77.5	60.5	74.0	**78.8**	75.1	69.5
Sensitivity	75.9	72.4	65.5	82.8	65.5	75.9	96.6	72.4	72.4
Specificity	77.8	77.8	77.8	72.2	55.6	72.2	61.1	77.8	66.7
F1 score	80.0	77.8	73.1	82.8	67.9	78.6	87.5	77.8	75.0
AUC	75.3	75.9	69.2	79.5	54.8	73.9	68.6	72.4	67.2

SS: Single-task; DAn: Walking while naming animals; DS1: Serial subtraction by 1s; DS7: Serial subtraction by 7s; F: Walking Fast; LR: Logistic Regression; RF: Random Forest; SVM: Support Vector Machine; RFE: Recursive feature elimination; AUC: Area under the curve. [a)] using dual-task costs; [b)] using dual-task costs (or capacity indexes) and SS metrics

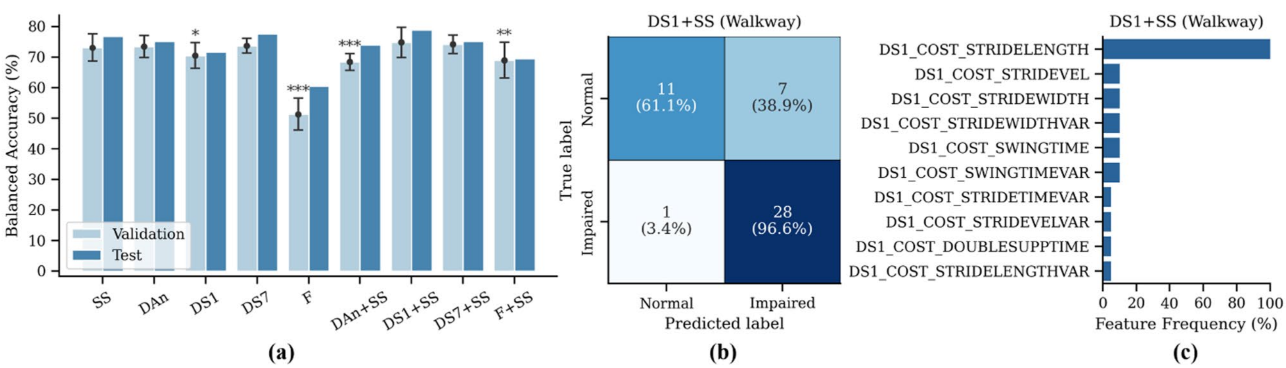

Fig. 3 Performance using features extracted from the walkway. (**a**) Performance on validation and test sets, where vertical lines indicate standard deviation of the 20 inner CVs. The statistical analysis compares the best-performing experiment (DS1 + SS) with the results achieved in the other conditions, using Paired Samples *T*-tests or Wilcoxon signed-rank tests with Bonferroni correction, where $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$ indicate statistical significance. (**b**) Confusion matrix of the DS1 + SS condition. (**c**) Feature selection frequency for the DS1 + SS condition

Figure 2b shows the confusion matrix of the DS1 + SS condition.

Classification using walkway metrics

The results achieved using the metrics extracted from the electronic walkway are shown in Table 4. As with the accelerometer, the best performing condition was the combination DS1 + SS, with 78.8% of balanced accuracy, 96.6% of sensitivity, 61.1% of specificity, and 68.6% of AUC. When using features from the gait test alone, DS7 (with 77.5% of balanced accuracy) revealed the highest performance in comparison with SS, DAn, DS1, and F conditions. Balanced accuracies spanned from 60.5% (with F) to 78.8% (with DS1 + SS), being test scores higher than its equivalents with the accelerometer.

Although the electronic walkway provided an additional number of gait metrics than the accelerometer, the number of features used by machine learning models was not always higher. The SS condition, for instance, used only 2.0 ± 2.2 features, whereas in the case of the accelerometer an average number of 3.0 ± 0.0 features were used in this condition. However, the features included were different. In the SS condition, for instance, double support time was selected for all the 20 outer training folds and stride width variability was selected 5 times (results not shown). Double support time was one of the most relevant features of the electronic walkway in several conditions where feature selection has been applied, including, SS, DAn, DS1, F, DS7 + SS, and F + SS. In the case of the DS1 + SS condition, the best balanced accuracy of 78.8% was achieved with 1.7 ± 2.1 features, being stride length cost one of the most relevant features, selected in all folds of the outer training loop (Fig. 3c). Only 1 participant in the impaired group was misclassified using DS1 + SS features (Fig. 3b), leading to a very high sensitivity of 96.6%.

Compared to the DS1 + SS condition, only DS1, F, DAn + SS, and F + SS conditions revealed significantly lower performances (Fig. 3a). The performance of the DS1 + SS was superior to the SS in validation and test folds, although the differences in validation performances were not statistically significant.

A Paired Samples *T*-test revealed that the performance of the machine learning models in the DS1 + SS condition using walkway features was significantly higher than the performance of the DS1 + SS condition using the features extracted from the accelerometer ($T = -4.446$, $p < 0.001$).

Table 5 Performance using MoCA and a combination with gait

Classifier	MoCA <26	MoCA <27	Majority voting
Bal. accuracy	69.9	76.8	82.7
Sensitivity	62.1	75.9	93.1
Specificity	77.8	77.8	72.2
F1 score	70.6	80.0	88.5

MoCA: Montreal Cognitive Assessment

Performance using MoCA

Table 5 and Fig. 4 summarize the results of the classification when a single cut-point was employed to MoCA. Using the traditional cut-point of 26, the MoCA's sensitivity was 62.1% and the specificity was 77.8% in the present sample (Table 5). The ROC analysis (ROC curve shown in Fig. 4a) indicated that in the present sample the cutoff of 27 would optimize sensitivity and specificity, which was according to the results obtained in [27]. The cutoff of 27 yielded a sensitivity of 75.9%, a specificity of 77.8%, and an AUC of 66.0%.

Since the DS1 + SS condition measured by the walkway yielded the best classification performance when using gait metrics (Table 4), and because the DS1 + SS condition requires a SS test in order to measure the dual-task costs, we combined the classification results using a majority voting approach that considered the classifications provided by the SS condition, the DS1 + SS condition and MoCA<27. The majority voting resulted in an improved balanced accuracy of 82.7%, a sensitivity of 93.1%, and a specificity of 72.2%. The resulting confusion matrix is shown in Fig. 4b.

An exact McNemar's Test showed that there were no statistically significant differences in the disagreements between the classifications provided by MoCA<27 and the classifications provided by the machine learning models using the best performing condition of DS1 + SS (walkway), with $p = 0.549$. The observed differences in classification results were also not statistically significant when the traditional threshold (<26) was employed, with $p = 0.092$. Although the majority voting scheme revealed the highest balanced accuracy overall, the observed differences in classification results were not significantly different from DS1 + SS ($p = 1.0$), SS ($p = 0.289$), or MoCA<27 ($p = 0.219$), but differed significantly from MoCA<26 ($p < 0.05$).

Discussion

This study was the first to investigate the use of machine learning and gait characteristics for detecting CI in post-stroke patients using a gold-standard neuropsychological battery as a reference. We used a nested cross-validation approach to evaluate the performance of multiple classifiers and feature selection methods on different walking conditions, including single-task (SS), dual-task (DAn, DS1, and DS7), and fast walking (F). Moreover, we tested the inclusion of dual-task costs and capacity indexes resulting from the combination of SS with DAn, DS1, DS7, or F. In line with our initial hypothesis, gait characteristics were able to distinguish between post-stroke individuals with and without CI. Moreover, a comprehensive set of gait metrics and dual-task costs allowed better differentiation between the two groups, with classification performances that were comparable to or even superior to those achieved with MoCA. Combining gait metrics with MoCA improved overall classification performance.

Performance using gait metrics

The machine learning experiment leading to the best classification overall was achieved using DS1 + SS gait metrics extracted from the electronic walkway. Although

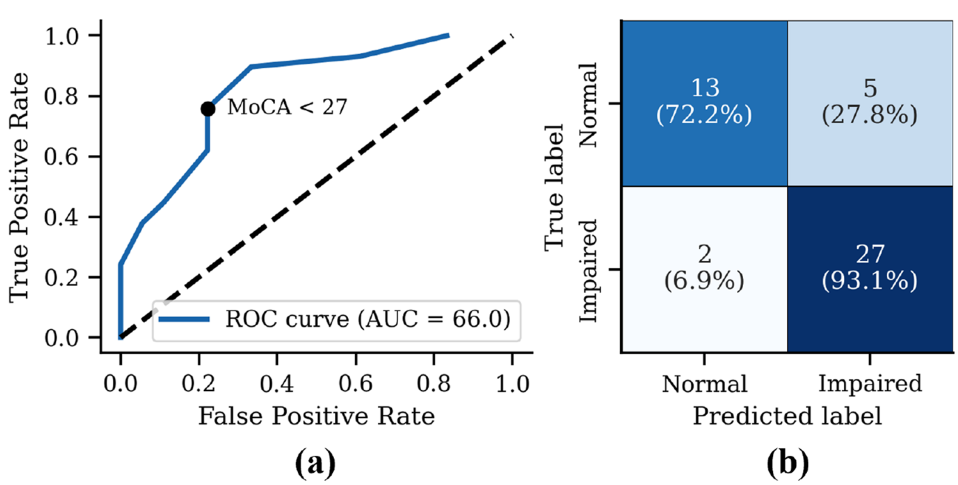

Fig. 4 Classification performance using MoCA and a majority voting approach. (**a**) ROC analysis used to determine the optimal MoCA cut-point and (**b**) Confusion matrix resulting from the majority voting approach, which combined the SS condition, the DS1 + SS condition (walkway) and MoCA<27. AUC: Area under the curve; ROC: receiver operating characteristic

disagreements in classification results were not significantly different from MoCA (using cutoffs of <26 or <27), the balanced accuracy achieved with gait metrics was higher (78.8% versus 69.9% or 76.8%, respectively), which is a positive indication that gait analysis can be a useful tool for the screening of CI in post-stroke patients. Compared to MoCA, gait tests are simpler and allow more objective and faster evaluations, which justifies their relevance in the context of cognitive screening [18].

The DS1 + SS condition was consistently identified as the best-performing condition, either using the accelerometer or the walkway. However, in both cases, the differences between the DS1 + SS condition and the SS condition were not statistically significant. This suggests that even a simple single-task gait analysis could provide acceptable classification results, without the need for participants to complete additional tests. Despite this, the sensitivity obtained in the DS1 + SS condition using the electronic walkway was much higher at 96.6%. This suggests that the DS1 + SS condition should be a preferred choice for cognitive screening, where the main goal is to detect all potential cases for further assessment [8].

The lower classification performances obtained with the accelerometer may be in part due to the lower level of detail it provided compared to the electronic walkway. While the accelerometer could only provide general gait parameters, like, velocity, cadence, and total steps, the electronic walkway provided additional spatio-temporal metrics that characterized the strides (e.g., length, speed, width) and gait phases (e.g. swing and double support time). We should note that the limitations of the accelerometer in this regard are not due to the device itself, but rather to the algorithms used to process its data [48–50]. It's also worth noting that in the analysis of the descriptive statistics, the electronic walkway detected significant differences in gait speeds between the two groups in more tests than the accelerometer (Table 2). This may suggest that there are differences in the performance of the two technologies. In fact, different algorithms can be used to process data from inertial sensors, and the inclusion of a gyroscope can often lead to improved results [48–50]. Therefore, the choice of algorithm and sensors used may have an impact on the results obtained. Thus, a possible lack of instrument accuracy (and/or precision) could partially justify the obtained results. Yet, including more detailed evaluations of gait seems to be beneficial, which is according to previous results on the detection of CI. According to previous studies, gait speed may be associated with several adverse health outcomes [51, 52], and therefore, it is not specific for the detection of CI [53, 54]. Since no other pathologies were included in our study, these results could not be confirmed.

Previous studies showed that different cognitive tasks can impact walking in different ways. In general, more cognitively demanding tasks tend to lead to greater interference effects, although the results are not always consistent [13]. The serial subtraction by sevens is particularly cognitively demanding and has been shown to be the most sensitive in distinguishing patients with CI in non-stroke populations [13]. In this study, the classification performance achieved with the DS7 and DS7 + SS conditions was significantly lower than the DS1 + SS condition when using the accelerometer (Fig. 2a), but did not differ from the DS1 + SS condition when using the walkway (Fig. 3a). However, the DS1 + SS condition led to higher performances in both cases, indicating better discrimination ability. It's worth noting that post-stroke patients often have other disabilities, particularly physical ones [9], and the mechanisms underlying motor-cognitive interferences may be different in this population compared to patients with other neurocognitive disorders such as Alzheimer's disease. Therefore, further research is needed to understand the effects of different cognitive tasks on post-stroke patients.

Previous research has also shown that combining multiple gait metrics can provide better discrimination ability than using a single gait metric [47, 55]. However, some gait metrics may be highly related and their simultaneous inclusion may not provide additional discrimination ability [56]. In this study, we used feature selection methods to identify the most relevant features based on different criteria. In the best-performing condition (DS1 + SS using the walkway), features were selected based on feature importance as provided by the Random Forest. In this experiment, only dual-task costs were selected for classification. According to previous studies, dual-task costs represent only the effect of adding the cognitive task and are not influenced by individual baseline characteristics (e.g., leg length) that could otherwise be considered covariates [14, 57]. For this reason, dual-task costs constitute an interesting evaluation metric for the detection of cognitive disorders.

The performances achieved in this study are at the level of the performances reported in the literature for the detection of CI in non-stroke patients. For instance, using dual-task gait data from an electronic walkway, Boettcher et al. [20] reported sensitivities of 82.0% and specificities of 67.7% on the detection of mild CI. Also with an electronic walkway, Ghoraani et al. [21] differentiated cognitively impaired groups from the healthy subjects with an accuracy of 86.0% and an F1 score of 88.0%. Using a detailed evaluation of gait provided by an inertial sensor-based gait analysis solution, Shahzad et al. [22] could achieve classification accuracies of 70.0% and sensitivities of 83.3% for the detection of mild CI, which proves the capacity of inertial sensor-based solutions.

Performance using MoCA

Similarly to the study by Zaidi et al. [27], the threshold of 27 in MoCA was the one achieving the best combination of sensitivity and specificity (Table 5). Although this study used the same dataset, they included all 161 participants and employed different neuropsychological criteria to characterize CI. In [27] the threshold of 27 resulted in a maximized sensitivity of 79% and specificity of 67% in detecting CI post-stroke. As in our study, increasing the cutoff from 26 to 27 allowed an increase in sensitivity.

To increase specificity, some studies recommend lowering the threshold of MoCA [8, 10]. A pooled analysis showed that MoCA scores below 26 are able to detect post-stroke CI with a sensitivity of 95% and a specificity of 45%, but lowering the threshold to 22 results in a lower sensitivity of 84% and a better specificity of 78% [8]. The differences in sensitivity and specificity reported by the different studies are mostly justified by the different neuropsychological criteria used to characterize CI [12].

Although disagreements in classification results obtained with MoCA<27 and with the best-performing gait condition did not differ significantly when compared to the majority voting approach, the combination of these two modalities resulted in an overall improvement in classification performance (Table 5). Compared to MoCA<26, the disagreements in classifications were significantly different, with the majority voting scheme achieving better sensitivity (93.1% versus 62.1%) for detecting CI. The sensitivity also improved compared to MoCA<27 (93.1% versus 75.9%), making the majority voting approach more suitable for screening purposes [8]. Although its sensitivity was lower than that of the DS1 + SS condition (93.1% versus 96.6%), it still remained very high, with a large improvement in specificity (72.2% versus 61.1%), which should favor its choice. The majority voting approach has received much attention in the literature for its simplicity and good performance on real data and has been applied in various fields [43]. To the best of our knowledge, this is the first study to propose the combination of MoCA with gait metrics for classifying cognitive status in post-stroke patients.

Study limitations

Several factors constitute the limitations of our study. First, the sample size was relatively small, which may affect the robustness of the models and the generalizability of our findings. Second, the criteria for classifying participants as impaired or cognitively normal based on neuropsychological assessments are not universally accepted and may vary depending on the cognitive domains evaluated, the tests used, their grouping criteria, the cutoffs demarcating impairment (most commonly 1 SD, 1.5 SD, or 2 SD), and the number of tests required to declare a deficit in a certain domain [12, 34]. In our study, we employed the comprehensive criteria recommended by Jak et al. [34] and the cognitive domains proposed by Dilliott et al. [33], which resulted in 61.7% of the patients meeting the criteria for CI. This frequency is above the pooled prevalence rates reported in the literature for the first year after stroke [12], but it is important to note that different criteria could lead to different results. Third, the participants in this study may have higher functioning levels than those typically seen in the general clinic, as those with severe functional disabilities were excluded from the study. Moreover, only individuals with MoCA>18 were included, which reduced the range of cognitive disorders associated with stroke that could be tested in this study. However, it is worth noting that our approach was able to detect CI in this group, which included more subtle symptoms that are usually harder to detect. Fourth, we did not evaluate the impact of the different study sites on the results. Although efforts were made to standardize assessments and ensure data quality across sites [24, 32, 36], differences could still be observed due to discrepancies in acquisition. This effect should be explored in future research. Finally, it is important to note that we did not evaluate the impact of the walking task on cognitive performance, although Plummer et al. [58] argue that the correct interpretation of the dual-task interference requires an objective evaluation of both tasks. Motor-cognitive interference may be observed in just one of the tasks, or in both simultaneously, and, for this reason, the performance on both tasks should be taken into account [58]. Additionally, it is worth noting that the order of the gait tests was maintained for all participants, which could potentially affect the results.

Future work

Further research is needed to determine the feasibility and effectiveness of gait assessments and machine learning models in detecting CI in post-stroke patients within a clinical setting. To achieve this, it would be necessary to include a larger and more diverse group of participants, including those with lower functioning levels and a wider range of cognitive abilities. Additionally, we should explore the detection of CI in the context of chronic CVD without a previous stroke [9].

Utilizing additional gait metrics as measured by inertial sensors could also be a valuable area of investigation. Inertial sensors have gained popularity in gait assessment due to their flexibility, low cost, and performance [59], and have the potential to extract gait metrics that have not yet been explored in this context [48, 49]. As suggested by Jung et al. [23], using sequential gait parameters extracted from inertial sensors could improve classification results and should, thus, be explored in the future. Furthermore, it is important to include an analysis of the performance on the cognitive task to provide a complete assessment of motor-cognitive interference in these patients [58].

Conclusion

Cognitive impairment (CI) after stroke is a common but often undiagnosed condition. To better manage cognitive disorders resulting from cardiovascular disease, it is necessary to develop new techniques to improve the screening of CI. In this study, we examined the use of gait, dual tasks, and machine learning to detect CI in post-stroke patients. Our results showed that a machine learning approach combined with dual-task gait metrics can effectively differentiate impaired and cognitively normal groups, with a performance that is not inferior to commonly used screening tests like MoCA. We also found that when using gait to assess cognitive status, it is important to include gait features that evaluate individual strides and gait phases, as opposed to just overall gait metrics like speed, cadence, or number of steps, as these result in poorer classification results. Using a majority voting approach that combines MoCA and dual-task gait metrics resulted in improved classification results, with better combinations of sensitivity and specificity for the detection CI. This study demonstrates the potential of machine learning and gait assessments as objective tools for cognitive screening, offering a good alternative or complement to MoCA in identifying CI in post-stroke patients. Further research is required to evaluate the practical and effective performance of this approach in a clinical setting.

Abbreviations

CVD	Cerebrovascular disease
VCI	Vascular cognitive impairment
CI	Cognitive impairment
MoCA	Montreal Cognitive Assessment
ONDRI	Ontario Neurodegenerative Disease Research Initiative
SS	Single-task
DAn	Walking while naming animals
DS1	Serial subtraction by 1s
DS7	Serial subtraction by 7s
F	Walking Fast
LR	Logistic Regression
RF	Random Forest
SVM	Support Vector Machine
KNN	*k*-Nearest neighbors
RFE	Recursive feature elimination
SMOTE	Synthetic Minority Oversampling Technique
ROC	Receiver operating characteristic
AUC	Area under the curve
CV	Cross-validation
SD	Standard deviation
DTC	Dual-task cost

Acknowledgements

We would like to acknowledge the individuals and organizations that have made Data used for this research available including the Ontario Neurodegenerative Disease Research Initiative (PI: Michael Strong, Mario Masellis, Douglas Munoz, and Richard Swartz), the Ontario Brain Institute, the Brain-CODE platform, the Government of Ontario, and the Temerty Family Foundation.

Authors' Contributions

VG formulated the research question under the lead of IS and MVC. VG conducted data processing, analysis, and interpretation of the results, and produced the first version of the manuscript. All authors revised and approved the final manuscript.

Funding

This work was supported in part by the project Smart-Health-4-All (POCI-01-0247-FEDER-046115), co-funded by Portugal 2020, framed under the COMPETE 2020 (Operational Program Competitiveness and Internationalization) and European Regional Development Fund (ERDF) from European Union (EU). It was also supported by the project ConnectedHealth (n.º 46858), funded by Competitiveness and Internationalisation Operational Programme (POCI) and Lisbon Regional Operational Programme (LISBOA 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Data Availability

All baseline data collected during the ONDRI foundational study, including all data supporting the findings of this study, are available on request from the Ontario Brain Institute (OBI) (details at http://www.ondri.ca).

Declarations

Ethics approval and consent to participate

The current study received approval from the Ethics Committee of the Faculty of Medicine of the University of Porto (51/CEFMUP/2022). The data used in this study were de-identified and were available on request from the Ontario Brain Institute (OBI). Access to the ONDRI Foundational Study Data was granted after approval by Brain-CODE's Data Access Committee, being ethics approval and consent to participate covered by the original ONDRI study (details at http://www.ondri.ca). The study is in accordance with the Declaration of Helsinki. All participants provided informed written consent.

Consent for Publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

Received: 29 March 2023 / Accepted: 18 March 2025
Published online: 01 April 2025

References

1. Dichgans M, Leys D. Vascular cognitive impairment. Circ Res. 120(3):573–91.
2. Aam S, Einstad MS, Munthe-Kaas R, Lydersen S, Ihle-Hansen H, Knapskog A-B, Ellekjær H, Seljeseth Y, Saltvedt I. Post-stroke cognitive impairment—impact of follow-up time and stroke subtype on severity and cognitive profile: the nor-COAST study. Front Neurol. 2020;11:699. https://doi.org/10.3389/fneur.2020.00699
3. Melkas S, Jokinen H, Hietanen M, Erkinjuntti T. Poststroke cognitive impairment and dementia: prevalence, diagnosis, and treatment. Degener Neurol Neuromuscul Dis. 2014;21. https://doi.org/10.2147/DNND.S37353
4. Calabrese P, Sitek EJ, Korczyn AD, Dong Y, Manso-Calderón R, Sierra-Beltrán M, Skrzypkowska A, Stefanova E. The assessment of cognitive and behavioural disturbances in vascular cognitive impairment (VCI) — recommendations of an expert working group. Neurologia I Neurochirurgia Polska. 2021;55(4):333–45. https://doi.org/10.5603/PJNNS.a2021.0035
5. Al-Qazzaz N, Ali S, Ahmad SA, Islam S, Mohamad K. Cognitive impairment and memory dysfunction after a stroke diagnosis: a post-stroke memory assessment. Neuropsychiatr Dis Treat. 2014;1677. https://doi.org/10.2147/NDT.S67184
6. Kalaria RN, Maestre GE, Arizaga R, Friedland RP, Galasko D, Hall K, Luchsinger JA, Ogunniyi A, Perry EK, Potocnik F, Prince M, Stewart R, Wimo A, Zhang Z-X, Antuono P. Alzheimer's disease and vascular dementia in developing countries: prevalence, management, and risk factors. Lancet Neurol. 2008;7(9):812–26. https://doi.org/10.1016/S1474-4422(08)70169-8

7. Mellon L, Brewer L, Hall P, Horgan F, Williams D, Hickey A. Cognitive impairment six months after ischaemic stroke: a profile from the ASPIRE-S study. BMC Neurol. 2015;15(1):31. https://doi.org/10.1186/s12883-015-0288-2
8. Lees R, Selvarajah J, Fenton C, Pendlebury ST, Langhorne P, Stott DJ, Quinn TJ. Test accuracy of cognitive screening tests for diagnosis of dementia and multidomain cognitive impairment in stroke. Stroke. 2014;45(10):3008–18. https://doi.org/10.1161/STROKEAHA.114.005842
9. Verdelho A, Wardlaw J, Pavlovic A, Pantoni L, Godefroy O, Duering M, Charidimou A, Chabriat H, Biessels GJ: Cognitive impairment in patients with cerebrovascular disease: a white paper from the links between stroke ESO Dementia Committee. Eur Stroke J. 2021;6(1):5–17. https://doi.org/10.1177/23969873211000258
10. Mijajlović MD, Pavlovié A, Brainin M, Heiss W-D, Quinn TJ, Ihle-Hansen HB, Hermann DM, Assayag EB, Richard E, Thiel A, Kliper E, Shin Y-I, Kim Y-H, Choi S, Jung S, Lee Y-B, Sinanović O, Levine DA, Schlesinger I, Mead G, Milošević V, Leys D, Hagberg G, Ursin MH, Teusch IY, Prokopenko S, Mozheyko E, Bezdenezhnykh A, Matz K, Aleksić V, Muresanu D, Korczyn AD, Bornstein NM. Post-stroke dementia - a comprehensive review. BMC Med. 2017;15(1):11. https://doi.org/10.1186/s12916-017-0779-7
11. Barbay M, Diouf M, Rousse IM, Godefroy O. GRECOGVASC study group: systematic review and meta-analysis of prevalence in post-stroke neurocognitive disorders in hospital-based studies. Dementia Geriatric Cognit Disord. 2018;46(5–6):322–34. https://doi.org/10.1159/000492920
12. Sexton E, McLoughlin A, Williams DJ, Merriman NA, Donnelly N, Rohde D, Hickey A, Wren M-A, Bennett K. Systematic review and meta-analysis of the prevalence of cognitive impairment no dementia in the first year post-stroke. Eur Stroke J. 2019;4(2):160–71. https://doi.org/10.1177/2396987318825484
13. Yang Q, Tian C, Tseng B, Zhang B, Huang S, Jin S, Mao J. Gait change in dual task as a behavioral marker to detect mild cognitive impairment in elderly persons: a systematic review and meta-analysis. Arch Phys Med Rehabil. 2020;101(10):1813–21. https://doi.org/10.1016/j.apmr.2020.05.020
14. Lowe DA, MacAulay RK, Szeles DM, Milano NJ, Wagner MT. Dual-task gait assessment in a clinical sample: implications for improved detection of mild cognitive impairment. J Gerontol B. 2020;75(7):1372–81. https://doi.org/10.1093/geronb/gbz119
15. Montero-Odasso M, Oteng-Amoako A, Speechley M, Gopau IK, Beauchet O, Annweiler C, Muir-Hunter SW. The motor signature of mild cognitive impairment: results from the gait and brain study. J Gerontol A Biol Sci Med Sci. 2014;69(11):1415–21. https://doi.org/10.1093/gerona/glu155
16. Inasaridze K, Foley JA, Logie RH, Sala SD. Dual task impairments in vascular dementia. Behav. Neurol. 2010;22(1–2):45–52. https://doi.org/10.1155/2010/604919
17. Mateus M. Performance characterization of dual-task in community dwelling elderly people and mild vascular cognitive impairment. Master's thesis, Universidade de Lisboa, Faculdade de Medicina, Lisboa, Portugal. 2019. http://hdl.handle.net/10451/39415
18. Aoki K, Ngo TT, Mitsugami I, Okura F, Niwa M, Makihara Y, Yagi Y, Kazui H. Early detection of lower MMSE scores in elderly based on dual-task gait. IEEE Access. 2019;7:40085–94. https://doi.org/10.1109/ACCESS.2019.2906908
19. Teh S-K, Rawtaer I, Tan HP. Predictive accuracy of digital biomarker technologies for detection of mild cognitive impairment and pre-frailty amongst older adults: a systematic review and meta-analysis. IEEE J Biomed Health Inform. 2022;26(8):3638–48. https://doi.org/10.1109/JBHI.2022.3185798
20. Boettcher LN, Hssayeni M, Rosenfeld A, Tolea MI, Galvin JE, Ghoraani B. Dual-task gait assessment and machine learning for early-detection of cognitive decline. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020;3204–7. IEEE, Montreal, QC, Canada. https://doi.org/10.1109/EMBC44109.2020.9175955
21. Ghoraani B, Boettcher LN, Hssayeni MD, Rosenfeld A, Tolea MI, Galvin JE. Detection of mild cognitive impairment and Alzheimer's disease using dual-task gait assessments and machine learning. Biomed. Signal Process. Control. 2021;64:102249. https://doi.org/10.1016/j.bspc.2020.102249
22. Shahzad A, Dadlani A, Lee H, Kim K. Automated prescreening of mild cognitive impairment using shank-mounted inertial sensors based gait biomarkers. IEEE Access. 2022;10:15835–44. https://doi.org/10.1109/ACCESS.2022.3149100
23. Jung D, Kim J, Kim M, Won CW, Mun K-R. Classifying the risk of cognitive impairment using sequential gait characteristics and long short-term memory networks. IEEE J Biomed Health Inform. 2021;25(10)4029–40. https://doi.org/10.1109/JBHI.2021.3073372
24. Farhan SMK, Bartha R, Black SE, Corbett D, Finger E, Freedman M, Greenberg B, Grimes DA, Hegele RA, Hudson C, Kleinstiver PW, Lang AE, Masellis M, McIlroy WE, McLaughlin PM, Montero-Odasso M, Munoz DG, Munoz DP, Strother S, Swartz RH, Symons S, Tartaglia MC, Zinman L, Investigators ONDRI, Strong MJ: The ontario neurodegenerative disease research initiative (ONDRI). Can J Neurol Sci. 2017;44(2):196–202. https://doi.org/10.1017/cjn.2016.415
25. Sunderland KM, Beaton D, Arnott SR, Kleinstiver P, Kwan D, Lawrence-Dewar JM, Ramirez J, Tan B, Bartha R, Black SE, Borrie M, Brien D, Casaubon LK, Coe BC, Cornish B, Dilliott AA, Dowlatshahi D, Finger E, Fischer C, Frank A, Fraser J, Freedman M, Greenberg B, Grimes DA, Hassan A, Hatch W, Hegele RA, Hudson C, Jog M, Kumar S, Lang A, Levine B, Lou W, Mandzia J, Marras C, McIlroy W, Montero-Odasso M, Munoz DG, Munoz DP, Orange JB, Park DS, Pasternak SH, Faria FP, Rajji TK, Roberts AC, Robinson JF, Rogaeva E, Sahlas DJ, Saposnik G, Scott CJM, Seitz D, Shoesmith C, Steeves TDL, Strong MJ, Strother SC, Swartz RH, Symons S, Tang-Wai DF, Tartaglia, M.c.t, A.K.T, Zinman J, L.m, P.M.M, Binns M, M.A. The Ontario neurodegenerative disease research initiative. preprint, neurology (August 2020). https://doi.org/10.1101/2020.07.30.20165456
26. Haddad SMH, Scott CJM, Ozzoude M, Berezuk C, Holmes M, Adamo S, Ramirez J, Arnott SR, Nanayakkara ND, Binns M, Beaton D, Lou W, Sunderland K, Sujanthan S, Lawrence J, Kwan D, Tan B, Casaubon L, Mandzia J, Sahlas D, Saposnik G, Hassan A, Levine B, McLaughlin P, Orange JB, Roberts A, Troyer A, Black SE, Dowlatshahi D, Strother SC, Swartz RH, Symons S, Montero-Odasso M, Investigators O, Bartha R, Wilson D. Comparison of diffusion tensor imaging metrics in normal-appearing white matter to cerebrovascular lesions and correlation with cerebrovascular disease risk factors and severity. J Biomed Imaging. 2022;2022. https://doi.org/10.1155/2022/5860364
27. Zaidi K-B, Rich JB, Sunderland KM, Binns MA, Truong L, McLaughlin PM, Pugh B, Kwan D, Beaton D, Levine B, Sahlas DJ, Dowlatshahi D, Hassan A, Mandzia J, Troyer AK, Swartz RH. Methods for improving screening for vascular cognitive impairment using the montreal cognitive assessment. Can J Neurol Sci. 2020;47(6), 756–63. https://doi.org/10.1017/cjn.2020.121
28. Ontario Brain Institute: brain-CODE Portal (2022). https://www.braincode.ca/ Accessed 2022-07-05
29. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol. Psychiatry. 2003;54(5):573–83. https://doi.org/10.1016/S0006-3223(02)01866-8
30. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695–99. https://doi.org/10.1111/j.1532-5415.2005.53221.x
31. Abd Razak MA, Ahmad NA, Chan YY, Mohamad Kasim N, Yusof M, Abdul Ghani MKA, Omar M, Abd Aziz FA, Jamaluddin R. Validity of screening tools for dementia and mild cognitive impairment among the elderly in primary health care: a systematic review. Public Health. 2019;169:84–92. https://doi.org/10.1016/j.puhe.2019.01.001
32. McLaughlin PM, Sunderland KM, Beaton D, Binns MA, Kwan D, Levine B, Orange JB, Peltsch AJ, Roberts AC, Strother SC, Troyer AK. The quality assurance and quality control protocol for neuropsychological data collection and curation in the ontario neurodegenerative disease research initiative (ONDRI) study. Assessment. 2021;28(5):1267–86. https://doi.org/10.1177/1073191120913933
33. Dilliott AA, Sunderland KM, McLaughlin PM, Roberts AC, Evans EC, Abrahao A, Binns MA, Black SE, Borrie M, Casaubon LK, Dowlatshahi D, Finger E, Fischer CE, Frank A, Freedman M, Grimes D, Hassan A, Jog M, Kumar S, Kwan D, Lang AE, Mandzia J, Marras C, Masellis M, McIntyre AD, Pasternak S, Pollock BG, Rajji TK, Robinson JF, Rogaeva E, Sahlas DJ, Saposnik G, Sato C, Seitz D, Shoesmith C, Steeves T, Strother SC, Swartz RH, Tan B, Tang-Wai D, Tartaglia MC, Troyer AK, Turnbull J, Zinman L, Hegele RA. Association of apolipoprotein E variation with cognitive impairment across multiple neurodegenerative diagnoses. Neurobiol Aging. 2021;105:378–13789. https://doi.org/10.1016/j.neurobiolaging.2021.04.011
34. Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, Delis DC. Quantification of five neuropsychological approaches to defining mild cognitive impairment. Am J Geriatric Psychiatry. 2009;17(5):368–75. https://doi.org/10.1097/JGP.0b013e31819431d5
35. Nugroho H, Utama NP, Surendro K. Normalization and outlier removal in class center-based firefly algorithm for missing value imputation. J Big Data. 2021;8(1):129. https://doi.org/10.1186/s40537-021-00518-7
36. Investigators ONDRI, Montero-Odasso M, Pieruccini-Faria F, Bartha R, Black SE, Finger E, Freedman M, Greenberg B, Grimes DA, Hegele RA, Hudson C,

Kleinstiver PW, Lang AE, Masellis M, McLaughlin PM, Munoz DP, Strother S, Swartz RH, Symons S, Tartaglia MC, Zinman L, Strong MJ, McIlroy W. Motor phenotype in neurodegenerative disorders: gait and balance platform study design protocol for the ontario neurodegenerative research initiative (ONDRI). J Alzheimers Dis. 2017;59(2):707–21. https://doi.org/10.3233/JAD-170149
37. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res. 2010;11:2079–107
38. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Phys Rev E. 2004;69(6):066138. https://doi.org/10.1103/PhysRevE.69.066138
39. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324
40. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. BMC Bioinf. 2013;14(1):106. https://doi.org/10.1186/1471-2105-14-106
41. Pedregosa F, Varoquaux G, Gramfort A, Miche lV, Thirion B, Grise lO, Blonde lM, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30
42. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cut-points using two ROC based criteria. Am J Epidemiol. 2006;163(7):670–75. https://doi.org/10.1093/aje/kwj063
43. Narasimhamurthy A. Theoretical bounds of majority voting performance for a binary classification problem. IEEE Trans Pattern Anal Mach Intell. 2005;27(12):1988–95. https://doi.org/10.1109/TPAMI.2005.249
44. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt. 2013;34(5):502–08. https://doi.org/10.1111/opo.12131
45. Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput 1998;10(7):1895–923. https://doi.org/10.1162/089976698300017197
46. Seabold S, Perktold, J.. Statsmodels: econometric and statistical modeling with python. In: 9th Python in Science Conference. 2010.
47. De Cock A-M, Fransen E, Perkisas S, Verhoeven V, Beauchet O, Remmen R, Vandewoude M. Gait characteristics under different walking conditions: association with the presence of cognitive impairment in community-dwelling older people. PloS One. 2017;12(6):0178566. https://doi.org/10.1371/journal.pone.0178566
48. Guimarães V, Sousa I, Correia MV. Orientation-invariant spatio-temporal gait analysis using foot-worn inertial sensors. Sensors. 2021;21(11):3940. https://doi.org/10.3390/s21113940. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute
49. Guimarães V, Sousa I, Correia MV. A deep learning approach for foot trajectory estimation in gait analysis using inertial sensors. Sensors. 2021;21(22):7517.https://doi.org/10.3390/s21227517
50. Petraglia F, Scarcella L, Pedrazzi G, Brancato L, Puers R, Costantino C. Inertial sensors versus standard systems in gait analysis: a systematic review and meta-analysis. Eur J Phys Rehabil Med. 2019;55(2): 265–80. https://doi.org/10.23736/S1973-9087.18.05306-6
51. Kyrdalen IL, Thingstad P, Sandvik L, Ormstad H. Associations between gait speed and well-known fall risk factors among community-dwelling older adults. Physiother Res Int. 2019;24(1):1743. https://doi.org/10.1002/pri.1743
52. Jung H-W, Jang I-Y, Lee CK, Yu SS, Hwang JK, Jeon C, Lee YS, Lee E. Usual gait speed is associated with frailty status, institutionalization, and mortality in community-dwelling rural older adults: a longitudinal analysis of the aging study of pyeongchang rural area. Clin Interv Aging. 2018;13:1079–89. https://doi.org/10.2147/CIA.S166863
53. Kikkert LHJ, Vuillerme N, van Campen JP, Hortobágyi T, Lamoth CJ. Walking ability to predict future cognitive decline in old adults: a scoping review. Ageing Res Rev. 2016;27:1–14. https://doi.org/10.1016/j.arr.2016.02.001
54. Bahureksa L, Najafi B, Saleh A, Sabbagh M, Coon D, Mohler MJ, Schwenk M. The impact of mild cognitive impairment on gait and balance: a systematic review and meta-analysis of studies using instrumented assessment. Gerontology. 2017;63(1):67–83. https://doi.org/10.1159/000445831
55. Hao W, Zhao W, Kimura T, Ukawa S, Kadoya K, Kondo K, Tamakoshi A. Association of gait with global cognitive function and cognitive domains detected by MoCA-J among community-dwelling older adults: a cross-sectional study. BMC Geriatr. 2021;21(1):523. https://doi.org/10.1186/s12877-021-02467-5
56. Pieruccini-Faria F, Black SE, Masellis M, Smith EE, Almeida QJ, Li KZH, Bherer L, Camicioli R, Montero-Odasso M. 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. 2021;17(8):1317–28. https://doi.org/10.1002/alz.12298
57. Verghese J, Robbins M, Holtzer R, Zimmerman M, Wang C, Xue X, Lipton RB. Gait dysfunction in mild cognitive impairment syndromes. J Am Geriatr Soc. 2008;56(7):1244–1251. https://doi.org/10.1111/j.1532-5415.2008.01758.x
58. Plummer P, Eskes G. Measuring treatment effects on dual-task performance: a framework for research and clinical practice. Frontiers in Human Neuroscience 2015;9. https://doi.org/10.3389/fnhum.2015.00225
59. Buckley C, Alcock L, McArdle R, Rehman R, Del Din S, Mazzà C, Yarna lIA, Rochester L. The role of movement analysis in diagnosing and monitoring neurodegenerative conditions: insights from gait and postural control. Brain Sci. 2019;9(2):34. https://doi.org/10.3390/brainsci9020034

Publisher's Note