scientific reports

Check for updates

OPEN

Identifying key gait features in stroke patients using wearable inertial sensors and supervised and unsupervised machine learning

Paolo Brasiliano[1,8]✉, Amaranta Soledad Orejel-Bustos[1,2], Valeria Belluscio[1], Andrea Cereatti[3], Ugo Della Croce[4], Dante Trabassi[5], Francesca Salis[4], Marco Tramontano[6,7], Maria Gabriella Buzzi[2], Giuseppe Vannozzi[1] & Elena Bergamini[8]

Stroke is a major cause of motor disability, degrading walking and quality of life. Wearable gait analysis with magneto-inertial measurement units (MIMUs) can quantify post-stroke impairments. We used machine learning to identify discriminative gait features in stroke, coupling supervised feature selection with unsupervised clustering to improve interpretability and generalizability. Eighty-five stroke patients and 97 healthy controls completed 10-Meter Walk Tests while wearing five MIMUs. Feature selection spanned spatiotemporal, symmetry, stability, and smoothness metrics. K-nearest neighbors (KNN), support vector machines (SVM), and decision trees (TREE) were trained, validated, and tested iteratively across data splits; clustering then verified discriminative ability. Sequential backward feature selection retained nine features, yielding accuracies (healthy vs. patient) of 94.1% (KNN), 96.7% (SVM), and 89.1% (TREE). SVM generalized best. Unsupervised k-medoids with cosine distance confirmed discrimination, reaching 90% accuracy with only three features: stride speed, stance-phase coefficient of variation, and medio-lateral harmonic ratio. Results indicate that gait variability, trunk smoothness, and upper-body stability robustly characterize post-stroke dysfunctions. Notably, head-movement smoothness emerged as a novel, discriminative feature. This integrated framework shows how wearable sensors plus machine learning can support clinical gait analysis and rehabilitation planning. Future work should enable real-time deployment and broaden datasets to cover more clinical scenarios.

The term stroke is used to describe brain damage due to several different vascular causes[1]. Stroke is a global challenge that poses significant health and socioeconomic challenges for both the individual and society as a whole[2,3]. Indeed, after a stroke event, brain cells die, resulting in functional and cognitive impairments[2]. Among these, motor disabilities primarily limit patients' ability to accomplish activities of daily living[4], with walking impairments influencing participation, autonomy, and quality of life of patients[5,6]. With these premises, the improvement of gait impairments is one of the crucial aspects of post-stroke rehabilitation[5]. In this context, standard clinical evaluation can be integrated with instrumented gait analysis, aimed at obtaining objective indicators of walking performance to evaluate the progression of the pathology, define tailored treatments, and monitor the efficacy of the latter. In the past years, the context in which instrumented gait analysis took place (especially in the clinical practice) has shifted towards a more ecological one, thanks to the increasing development of wearable technologies and the development of computational methods[7–10], bringing several advantages over traditional laboratory-based assessments. In this context, magneto-inertial measurement units (MIMUs) are widely used to obtain features related to the quality of gait and pertain to different domains, like symmetry, smoothness or spatiotemporal features which are often considered during post-stroke clinical

[1]Department of Motor, Human and Health Sciences, University of Rome "Foro Italico", Rome 00135, Italy. [2]IRCCS Santa Lucia Foundation, Rome 00179, Italy. [3]Department of Electronics and Telecommunications, Politecnico di Torino, Turin 10129, Italy. [4]Department of Engineering, University of Sassari, Sassari 07100, Italy. [5]Department of Medico – Surgical Sciences and Biotechnologies, Sapienza University of Rome, Latina 04100, Italy. [6]Department of Biomedical and Neuromotor Sciences, Alma Mater University of Bologna, Bologna 40138, Italy. [7]Unit of Occupational Medicine, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna 40138, Italy. [8]Department of Management, Information and Production Engineering, University of Bergamo, Bologna 24044, Italy. ✉email: paolobrasilianores@gmail.com

evaluations[8]. Often, in MIMU-based gait studies on stroke, data are typically collected during straight walking using one or more MIMUs placed on the trunk and/or lower limbs to compute spatiotemporal, symmetry, variability, smoothness, and regularity features[8,11–21]. Nevertheless, at present, the scientific works on the topic present a variety of methodological approaches in terms of sensor locations and numbers, as well as data processing techniques (i.e., gait events identification, signal preprocessing and type of gait features estimated). Given the large number of alternative pipelines and methodological choices reported in the literature, producing a single concise yet fully representative synthesis of all approaches is inherently challenging. Therefore, comparing results across studies and identifying which gait domains are consistently altered after stroke remains challenging, further motivating the need for robust and generalizable feature-selection strategies.

In combination with the availability of new measurement tools, new data analysis techniques, like machine learning algorithms, have been fruitfully integrated to obtain more information from patients' data. These tools have found many applications in the stroke population, ranging from the discrimination between/among groups or patient categories (like people with stroke and healthy controls, patients with different pathologies, or patients with different levels of stroke severity)[11], as well as the recognition of different types of activities, classification of well/poorly executed tasks, and other applications[8,12]. Among them, one of particular interest for clinicians is the classification of pathological and healthy people based on gait features. Although it may seem an obvious categorization following a medical diagnosis, the understanding of a pathology is primarily based on the characterization of the differences when compared to a non-pathological condition. Nevertheless, to do so, the features that can optimally capture such differences must be identified. Indeed, when analyzing gait, a wide number of features may be measured, some of which may not carry useful information for the pathological population of interest. To overcome this issue, features selection techniques and machine learning algorithms may be used in combination to reduce the number of features with the aim of retaining the feature subset able to distinguish healthy from pathological people and, consequently, identify those gait domains that altogether characterize the pathology. In this framework, classification performance per se is often employed as an indicator of the quality of the selected features in distinguishing between healthy and pathological conditions.

Although some authors have explored this (or similar) approach[13–18,20,22–24], most of them applied feature selection methodologies as an intermediate step to improve the performance of a given machine learning model, thus not focusing on the generalizability of the feature selection approach (i.e., the stability of the selected subset in its discriminative performance when changing the classifier, the training sample, and the validation strategy). Indeed, the selected features depend on several factors, including the dataset/sample size, the feature selection technique, and the machine learning algorithm employed. Previous studies applying machine learning (as well as deep learning) algorithms on stroke patients' gait data were conducted on too limited sample sizes and often using only one classification algorithm, thus not considering the effect of these issues on the discriminative performance of the selected features. Testing the generalizability of the feature selection approach across different machine learning algorithms is thus of the utmost importance. To the authors' knowledge, among recent studies leveraging MIMUs and machine learning in patients with stroke[13–17,20,21,24,25], only one[24] sought to reduce the features space by employing a feature selection technique across multiple machine learning algorithms and analyzing the frequency of the selected features. This approach identifies features that most consistently enhance classification performance, thus highlighting characteristics of pathological gait patterns. Nevertheless, the sample size in that study was insufficient to provide a reliable representation of the investigated population. Furthermore, no previous study has evaluated whether features selected through supervised machine learning could reliably differentiate healthy and pathological individuals when applied to unsupervised clustering methods. In other words, the discriminative value of selected features has not been tested independently of the supervised algorithm's learning capabilities.

Therefore, this study aims to identify an optimal subset of gait features extracted using a set of MIMUs through a feature selection approach in combination with multiple machine learning algorithms. The goal was achieved by distinguishing individuals with stroke from healthy controls, thereby characterizing pathological gait patterns. To ensure generalizability, the analysis was iterated across various combinations of feature subsets and participant groups.

Methods

Participants

Eighty-five patients with stroke (PwS; sex: 45 M – 40 F; age: 57 ± 16 yrs; mass: 71 ± 12 kg; stature: 1.69 ± 0.09 m) and 97 healthy participants (HP; sex: 50 M – 47 F; age: 48 ± 12 yrs; mass: 70 ± 18 kg; stature: 1.67 ± 0.08 m) were enrolled in this study. The two groups were matched by age and sex. The study was conducted in accordance with the World Medical Association Declaration of Helsinki and was approved by the Ethics Committee of the Institute for Research and Healthcare Santa Lucia (with protocol number CE/AG4/PROG.383 – 11 and subsequent integrations). Healthy participants between the age of 18 and 80 years were considered eligible for the study if they did not report any condition or use of medication that could have affected their motor performance. Stroke patients (both in the sub-acute and in the chronic phase of the pathology) that were able to walk without any device or need physical assistance were included in the study (Functional Ambulation Classification[26] scale score ≥ 3). Exclusion criteria for this group were cognitive deficits affecting the capacity of patients to understand the task instructions (Mini Mental State Examination[27] > 4), severe unilateral spatial neglect, severe aphasia, and presence of neurological, orthopedic, or cardiac comorbidities. All participants were included in the study after providing their informed consent.

Experimental set-up

Data collection was performed in the gym of the Institute for research and Healthcare Santa Lucia, in Rome. Participants were asked to perform a 10-Meter Walk Test (10-MWT) at their self-selected speed along a straight

walkway while wearing comfortable shoes. At the beginning of each trial, participants were instructed to maintain an orthostatic posture for five seconds. Each participant performed a minimum of three trials. During the trials, participants were equipped with five synchronized MIMUs (OPAL, APDM wearable technologies, Portland, USA). The MIMUs included a triaxial accelerometer, gyroscope, and magnetometer with full scale ranges of ±6 g, ±1500 deg/s, and ±6 Gauss, respectively, with a sampling rate of 128 Hz. Three MIMUs were fixed to the upper body of the participants, on the forehead (FH) on the occipital cranium bone close to the lambdoid suture of the head, at the center of the sternum (ST), and at the lower back (LB) level, in correspondence of L4-L5 vertebrae. The last two MIMUs were placed laterally on the distal part of the tibiae, slightly above the lateral malleoli, and securely fixed with Velcro straps. Attention was paid to each MIMU fixation to minimize the relative movement between the MIMU and the underlying bones. The five-sensor configuration was selected to balance feasibility in clinical practice with the ability to compute all the feature domains considered in this work. In our pipeline, the two shank sensors were required for robust stride segmentation and lower-limb spatiotemporal/symmetry-related descriptors; the lower back sensor provided a trunk reference for global walking dynamics; and the sternum and head sensors enabled upper-body measures. A reduced setup (e.g., trunk-only) would have prevented reliable stride segmentation in pathological gait within our preprocessing strategy and would have excluded limb-related asymmetry descriptors, whereas adding further sensors would increase setup time and patient burden without proportionate gain for the specific aims of the present feature-selection analysis.

Signal preprocessing

Data preprocessing was performed through implementation of customized algorithms in the MATLAB® Software R2021b (The MathWorks Inc., MA, US).

First, a consistent reference frame was defined for all participants. During the static phase of the 10-MWT, a time-invariant transformation aligned each MIMU's local reference system to a frame based on the gravity vector. Afterwards, the time-invariant transformation was applied to the accelerometer and gyroscope data recorded during the dynamic phase of the test. Finally, gravity was removed from the component of the acceleration signal aligned with the vertical axis of the reference frame. As a result, all data were expressed relative to a reference frame that approximated the anterior-posterior (AP), medio-lateral (ML), and cranio-caudal (CC) anatomical axes[28]. Accelerometer data were filtered using a second order Butterworth low-pass filter with a cut-off frequency of 10 Hz, while gyroscope data were filtered using a second order Butterworth low-pass filter with a cut-off frequency of 6 Hz[29].

Gait events were identified from ML angular velocity recorded by the MIMUs placed on the shanks of the participants while walking speed and gait spatial features were calculated through forward and backward integration of shank data in combination with a complementary filter[30] and zero-velocity update procedure[31]. For each identified stride, features of upper body movement stability[32,33], symmetry[34], and smoothness[35] were calculated from FH, ST, and LB MIMU data. In addition, symmetry[36] and variability of gait spatiotemporal features were also calculated. For the sake of readability, in the following sections spatiotemporal gait features are not described in detail as they represent standard measurements in instrumented gait analysis. Nevertheless, a complete list of spatiotemporal features is provided. For information on the equations for calculating each feature refer to the work by Bertoli et al.[31].

Spatiotemporal features

- Stride frequency.
- Stride speed.
- Stride length.
- Stride duration.
- Stance speed.
- Stance length.
- Stance duration.
- Swing speed.
- Swing length.
- Swing duration.
- Double support duration.
- Single support duration.

Stability features

Root Mean Square (RMS) was calculated from LB, ST, and FH MIMUs acceleration signals over each stride as follows:

$$RMS = \sqrt{\frac{\sum_{i=1}^{n} x^2}{n}}$$

Where X represents the acceleration values and n the number of samples of the considered stride.

Coefficient of attenuation[33](COA) was calculated from LB, ST, and FH MIMUs acceleration signals over each stride as follow:

$$COA = \left(1 - \frac{RMS_{uppersegment}}{RMS_{lowersegment}}\right) \times 100$$

Precisely, COAs were calculated from LB to ST, from LB to FH, and from ST to FH.

Symmetry features

Improved Harmonic Ratio[34](IHR) was calculated from LB, ST, and FH MIMUs acceleration signals over each stride as follow:

$$iHR_n = \frac{\sum_{i=1}^{n} P_I^i}{\sum_{i=1}^{n} (P_I^i + P_E^i)}$$

Where P_I^i and P_E^i are the power of the intrinsic and extrinsic n considered harmonics.

The Symmetry Angle[36](SA) was calculated from the spatiotemporal features of the side-paired values of each stride, as follows:

$$SA = \frac{\left(45^\circ - arctanarctan\left(\frac{X_{left}}{X_{right}}\right)\right)}{90^\circ} \times 100$$

Where X_{left} and X_{right} are the features values for left and right strides, respectively.

Smoothness features

Log dimensionless jerk[35](LDLJ) was calculated from LB, ST, and FH MIMUs acceleration and angular velocities (LDLJA and LDLJW, respectively) signals over each stride as follows:

$$LDLJ = -lnln\left(\frac{t^2 - t^1}{\left(||x\,(t)||\right)} \cdot Ij\right)$$

With:

$$Ij = \int_{t^1}^{t^2} \left|\left|x'\,(t)\right|\right|^2 dt$$

where x represents the linear acceleration or angular velocity signals and t^1 and t^2 are the starting and ending instants of each stride.

Finally, variability of spatiotemporal features was estimated by calculating the Coefficient of Variation (CoV) of each feature.

The median value was calculated over the gait cycles for each trial and each feature. Afterwards, dataset was visually inspected to detect outliers and trials that were considered to present obvious instrumental errors were discarded. Finally, each participant's median value was calculated over the trials. Missing values were replaced with the group median.

Feature selection and validation

From the entire sample considered, two subgroups of 18 PwS and 20 HP (Unsupervised Test Groups) were selected and kept aside for further analysis. Afterwards, two subgroups of 67 PwS and 77 HP (Feature Selection Groups) were randomly identified and used for the first steps of the feature selection procedure. Precisely, within-groups distribution of each feature was tested using the Shapiro-Wilk test. Afterwards, according to data distribution, an independent sample t-test or a Mann-Whitney U test was used to identify those features which differed significantly between HP and PwS[14,37]. Only these features were considered for further analysis.

To limit multicollinearity of the dataset, Pearson's correlation coefficients between all the retained feature pairs were calculated[37] and analyzed according to the following procedure:

1. The total number of correlations with $r > 0.5$ was calculated for each feature.
2. The feature (F_{maxr}) that showed the highest number of over-threshold correlations was kept while those that showed over-threshold correlations with F_{maxr} were discarded.

The procedure was iterated until any $r > 0.5$ was found. If two features with the same number of over-threshold correlations were found (i.e., if two F_{maxr} were found), one was chosen according to the suggestion of physical therapists of the neurorehabilitation hospital.

Finally, from the Feature Selection Groups, ten subgroups were defined by randomly selecting 70% of the participants (48 PwS and 54 HP) for the training and validation sets, and the remaining 30% of the participants (20 PwS and 23 HP) for the test sets. Afterward, on these datasets, a sequential backward feature selection (SBS) was implemented. This procedure allows the reduction of the number of features while preserving the performance of the classifier and the interpretability of the results. Indeed, no new combination of features is created (such as in Principal Component Analysis, Linear Discriminant Analysis or other features extraction techniques); rather, only relevant features from the original dataset are kept, making the procedure usable and the results interpretable in clinical settings[37]. This approach is implemented by evaluating the performance of a classifier while changing the set of features as described:

1. The complete dataset with k features is defined as the starting point and tested.

2. All the possible combinations of *k-1* features are tested.
3. The subset of *k-1* features with the best classification performance is identified.
4. The subset of features identified in step 2.3 is used as the new starting point and the procedure is repeated from step 2.2.
5. The feature selection continues until a stop criterion is met.

In this instance, the SBS process was carried out until one single feature remained. The evaluated classifiers were the K-nearest neighbors (KNN), the Support Vector Machine (SVM), and the decision tree (TREE) algorithms. The algorithms were selected for their nonparametric approach, which does not require a-priori assumptions on the dataset. Moreover, deep-learning approaches were not included because the primary objective of this study was feature interpretability and the identification of a compact subset of clinically meaningful descriptors. Deep networks typically learn latent representations from raw signals that are not directly comparable to explicit feature-selection outcomes, and they often require larger datasets to generalize reliably. During the SBS procedure, hyperparameters tuning for each algorithm was performed using a Bayesian optimization approach. Detailed information on the hyperparameters tuning, for each algorithm is provided in supplementary material. The classifiers' performance during SBS was assessed using a 5-fold cross validation approach and measured by classification accuracy, i.e., the ability to correctly classify participants irrespective of their group.

As a result, during SBS, the three algorithms were trained and cross-validated on all ten subgroups randomly identified. Each time, feature values for the training and the validation sets were standardized using the mean and the standard deviation of the training set; afterwards, the combination of best subsets and best algorithm hyperparameters were identified according to the highest classification accuracy obtained. The best identified model (i.e., the best subset of features with the tuned hyperparameters) was then tested on the corresponding test set.

From the thirty SBS procedures carried out, only those features that were selected in at least two of the ten runs were retained. Afterwards, only the shared features between datasets and algorithms were kept. Such features were arranged in descending order according to their number of occurrences at the end of the SBS procedures. Subsequently, an unsupervised algorithm was tested, adding one by one the ordered selected features, both on the held-out test set generated in each of the ten SBS runs to evaluate the robustness of the unsupervised clustering to data partitioning, and on the Unsupervised Test Groups kept aside before data analysis and not considered during the features selection procedure to assess overall discriminative performance of the selected subset of features. In these instances, a k-medoids algorithm was implemented with the initial medoids identified using the Single Pass Seed Selection algorithm[38] to obtain a deterministic solution. K-medoids was implemented to produce two clusters. Distances between the clusters medoids and the data points were measured using four different distance metrics described in the following equations.

- Cosine distance (CO).

$$CO = 1 - \frac{A \cdot B}{\|A\|\|B\|}$$

Where $\boldsymbol{A}$ and $\boldsymbol{B}$ are the vectors defined by the features median value of the medoid and of each participant, respectively.

- Squared Euclidean distance (SqEU).

$$SqEU = \sum_{i=1}^{n} (A_i - B_i)^2$$

- City Block distance (CB).

$$CB = \sum_{i=1}^{n} |A_i - B_i|$$

- Euclidean distance (EU).

$$EU = \sum_{i=1}^{n} \sqrt{(A_i - B_i)^2}$$

Where $\boldsymbol{A_i}$ and $\boldsymbol{B_i}$ are i^{th} feature median value of the medoid and of each participant.

Being unsupervised, the algorithm produces two unlabeled clusters. To assign the group labels (i.e., PwS and HP) to the two clusters, the following procedure was implemented:

1. The medians of each feature for the stroke and healthy groups and the two identified clusters were calculated and arranged to form n-dimensional vectors.
2. The Euclidean distance between each of the two cluster vectors and each of the two group vectors was calculated.
3. The smallest distance was used to label the cluster according to the corresponding known group.
4. The other cluster was labeled by exclusion.

Except for the SBS procedures, in which only classification accuracy was measured, all the other classification performances also included the recall, precision, and F1-score. The procedure described is graphically shown in Fig. 1.

Results

Supervised classification results

Of the initial 79 features, 60 were retained after the *t*-test and 20 after the correlation analysis (detailed results are reported in Annex A and B). The mean (±SD) accuracy during SBS procedures over the ten runs for KNN, SVM, and TREE were 94.1% ± 1.6%, 96.7% ± 2.1%, and 89.1% ± 2.2%, respectively. The mean results of the classification on the test sets are reported in Table 1. Figure 2 shows the same performance indexes across each run performed.

Overall, SVM achieved the highest average accuracy and F1-score and, importantly, the highest recall, indicating a stronger capability to correctly identify participants with stroke. KNN showed a similar level of accuracy but a different balance between error types, with higher precision and lower recall compared with SVM, suggesting a more conservative behavior (fewer false positives at the expense of more false negatives). In contrast, the decision tree consistently yielded the lowest performance across all metrics and exhibited larger dispersion, particularly for recall and precision, pointing to reduced robustness across different data splits. These patterns are mirrored in Fig. 2, where the run-by-run bar plots make the variability across the ten test sets explicit: SVM maintains generally high recall and F1 across runs, KNN displays higher precision in several runs

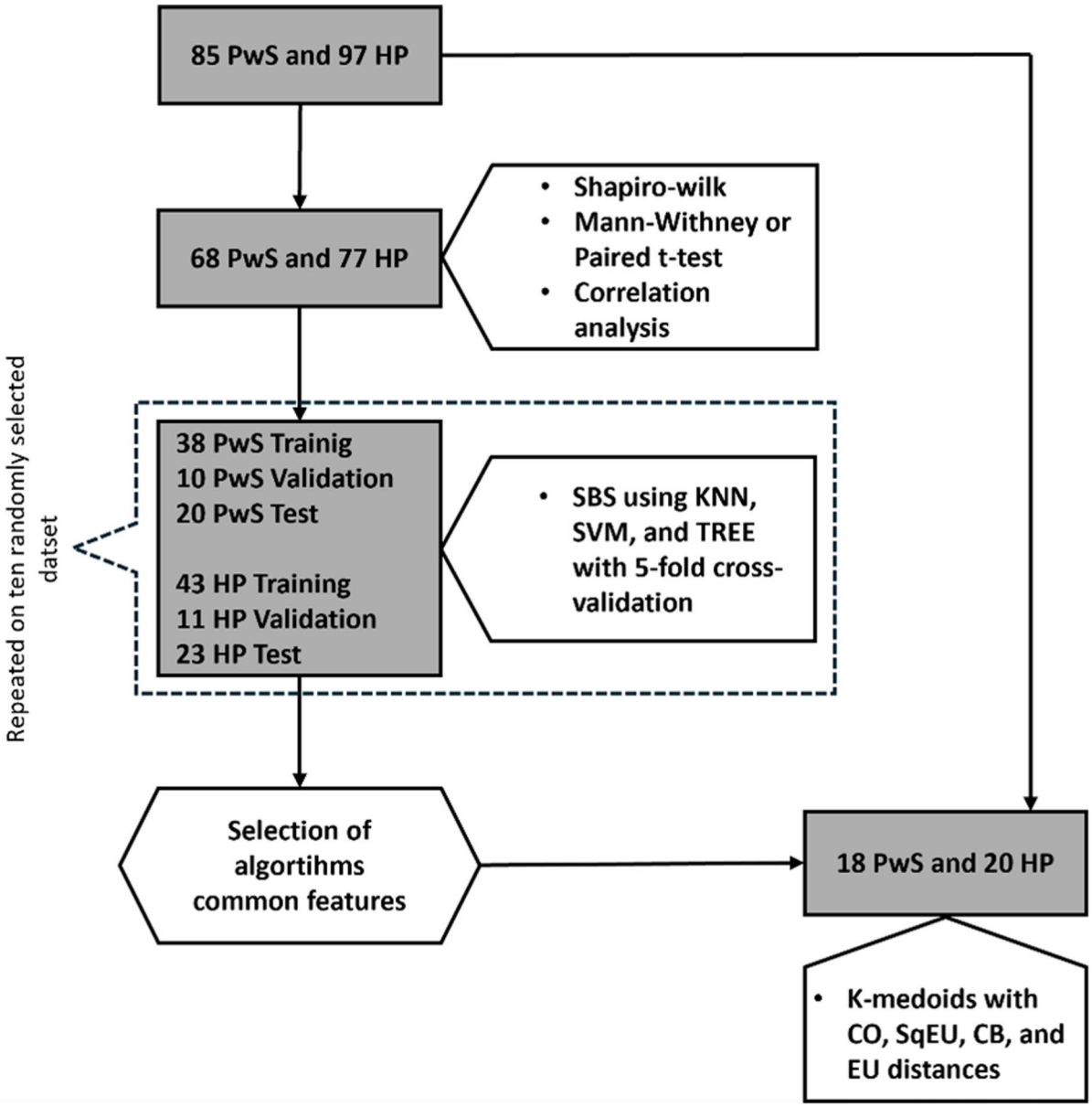

Fig. 1. Scheme of the steps followed during the data splitting and analysis presented in the work.

	Accuracy	Recall	Precision	F1-score
KNN	**88.1±5.7%**	**85±4.7%**	**89.4±8.3%**	**87.1±6%**
SVM	89.8±5.1%	91±5.7%	87.8±6.6%	89.2±5.3%
TREE	81.2±5.7%	78.5±9.1%	82.1±10.3%	79.6±5.5%

Table 1. Mean and standard deviation values of the performance indexes of the supervised algorithms on the test sets over the ten runs.

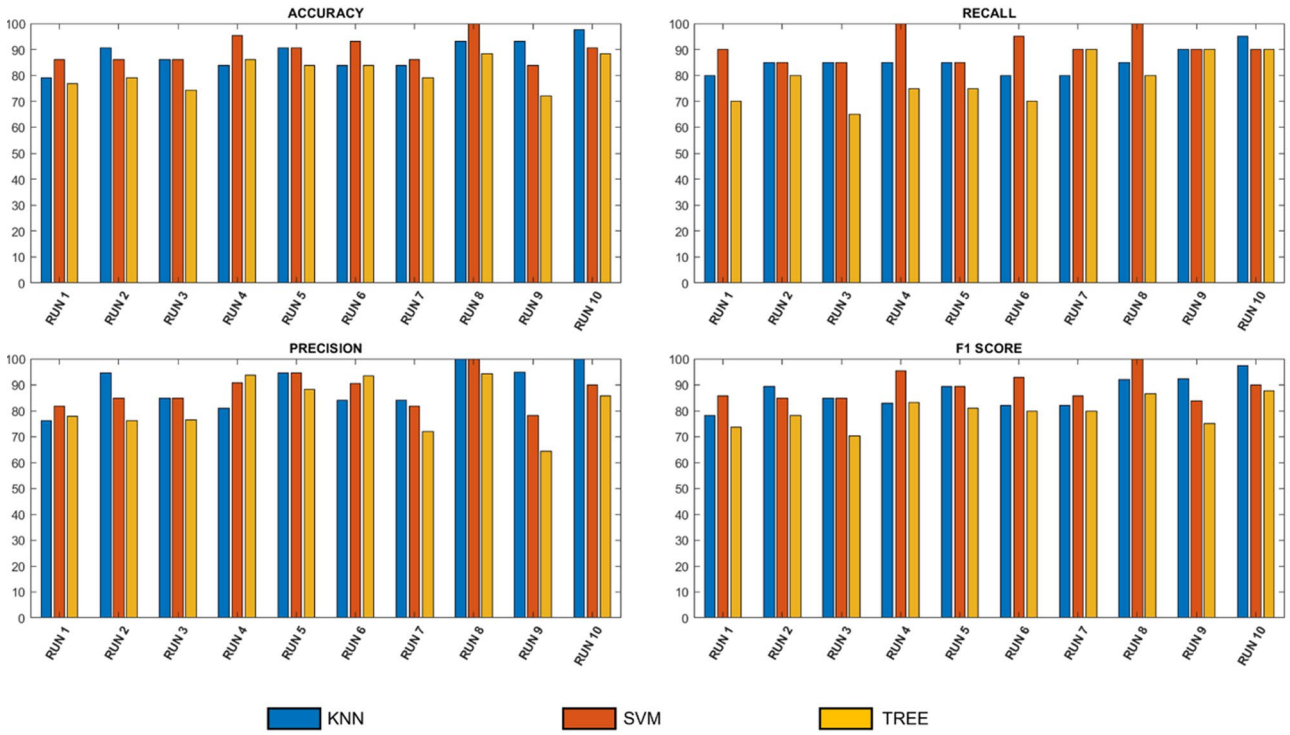

Fig. 2. Test-set performance across the ten SBS runs. Accuracy, recall, precision, and F1-score are reported for KNN, SVM, and TREE.

but with lower recall, and TREE shows both lower values and more pronounced fluctuations. Together, Table 1; Fig. 2 indicate that the supervised discrimination between healthy participants and stroke patients is generally high across the ten runs, with SVM providing the most stable and clinically favorable trade-off among the tested classifiers.

Feature selection results

The features used in the SBS procedure, together with their relative number of occurrences in total and for each algorithm, are shown in Fig. 3. The heatmap allows a direct comparison of selection frequencies across algorithms, highlighting both features that were repeatedly retained regardless of the classifier (i.e., stable "common" features) and features that were preferentially selected by only one algorithm (i.e., classifier-dependent selections). In addition, the final column provides the overall occurrence of each feature across the three classifiers, offering an immediate view of which descriptors were most consistently preserved by SBS.

As shown in the figure, a limited subset of features emerges with high occurrence values and is emphasized in bold, indicating the descriptors carried forward to the subsequent unsupervised clustering analysis. Specifically, the nine retained features were:

- Improved Harmonic Ratio in the medio-lateral direction measured at the lower back level (symmetry domain).
- Stance phase duration Coefficient of Variation (CoV) (variability domain).
- Stride speed (spatiotemporal domain).
- Swing phase duration (spatiotemporal domain).
- Head log-dimensionless jerk from acceleration signal in the medio-lateral direction (smoothness domain).
- Head log-dimensionless jerk from acceleration signal in the anterior-posterior direction (smoothness domain).
- Coefficient of attenuation between lower back and head in the medio-lateral direction (stability domain).

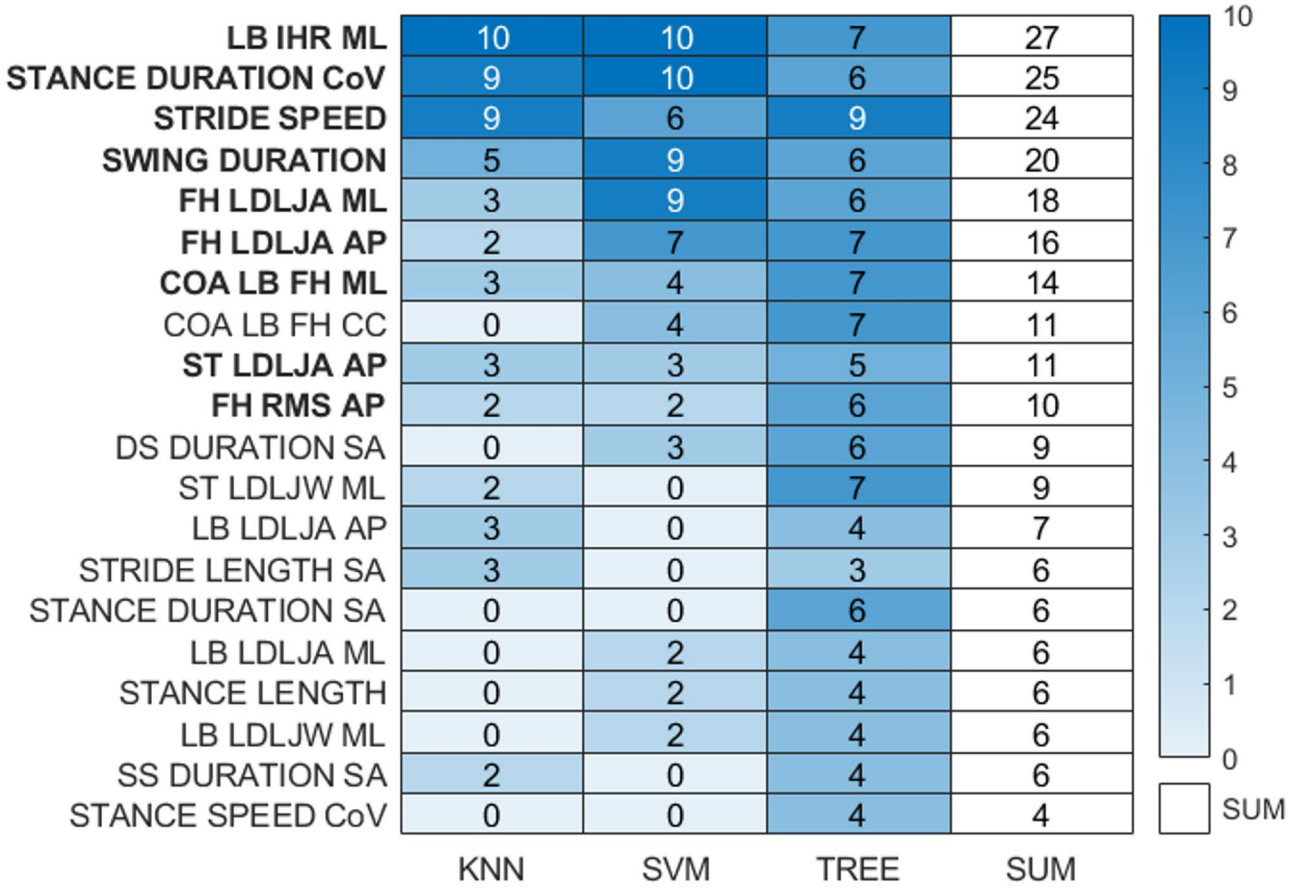

Fig. 3. Heatmap of feature-selection occurrences across the ten SBS runs for KNN, SVM, and TREE. The SUM column reports the overall occurrence of each feature across classifiers. Features in bold were retained for the subsequent analysis. For visualization purposes, some occurrence values are shown in white font to improve readability against darker cells. LB, ST, and FH: Lower Back, Sternum, and Forehead, respectively. AP, ML, and CC: Anterior–posterior, medio-lateral, and cranio-caudal, respectively. IHR: Improved Harmonic Ratio. CoV: Coefficient of Variation. LDLJ: Logarithmic dimensionless jerk from acceleration (LDLJA) or angular velocity (LDLJW) data. COA: Coefficient of attenuation.

- Sternum log-dimensionless jerk from acceleration signal in the anterior-posterior direction (smoothness domain).
- Head root mean square from acceleration signal in the anterior-posterior direction (stability domain).

Together, these features span spatiotemporal descriptors, variability, and upper-body gait quality measures, suggesting that discrimination between stroke and healthy gait is supported by both lower-limb timing/speed alterations and upper-body control characteristics. Overall, the distribution of occurrences indicates that, despite some algorithm-specific differences, several descriptors are robustly selected across classifiers and repetitions, supporting their relevance for distinguishing healthy participants from stroke patients.

Unsupervised clustering results

Unsupervised clustering was used to test whether the SBS-derived feature subset preserves discriminative structure independently of supervised learning. We report clustering repeated across the test sets defined during SBS procedures to quantify stability against split-to-split variability and SBS-induced changes in feature ordering (Fig. 4), and the primary unsupervised validation on the Unsupervised Test Groups (Fig. 5). More in detail, Fig. 4 shows the mean (with relative standard deviation) of the performance indexes across the ten runs as each common feature is added incrementally, ordered by number of occurrences. The best performance (90.2% ± 5.5%, 87% ± 5.8%, 89.2% ± 6% for accuracy, recall, and F1-score, respectively) on the test sets was obtained when only three features were used, namely the improved harmonic ratio on the medio-lateral direction, the coefficient of variation of the stance phase, and the stride speed. Overall, performance typically improves rapidly when the first features are added and then tends to plateau or fluctuate as additional descriptors are included, with larger variability observed for some distance metrics at intermediate steps, especially in recall- and F1-related trends.

Similarly, Fig. 5 presents the performance indexes for the Unsupervised Test Groups, showing the incremental addition of each common feature, providing the primary unsupervised generalization check. The overall pattern is consistent with the ten-run results: performance is already high with a reduced subset of features and does not

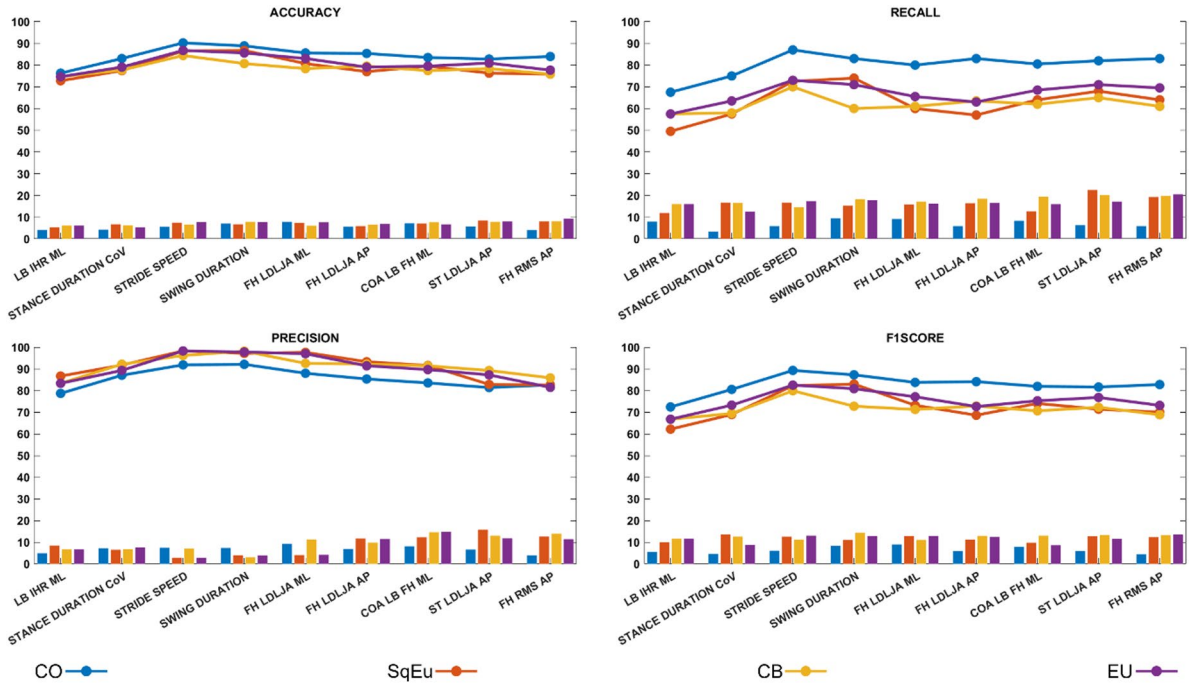

Fig. 4. Unsupervised clustering performance across the ten SBS test sets as features are added one by one, ordered by their total SBS occurrence. Accuracy, recall, precision, and F1-score are shown for k-medoids using four distance metrics (CO, SqEU, CB, and EU). Points represent the mean across the ten runs and error bars indicate the standard deviation.

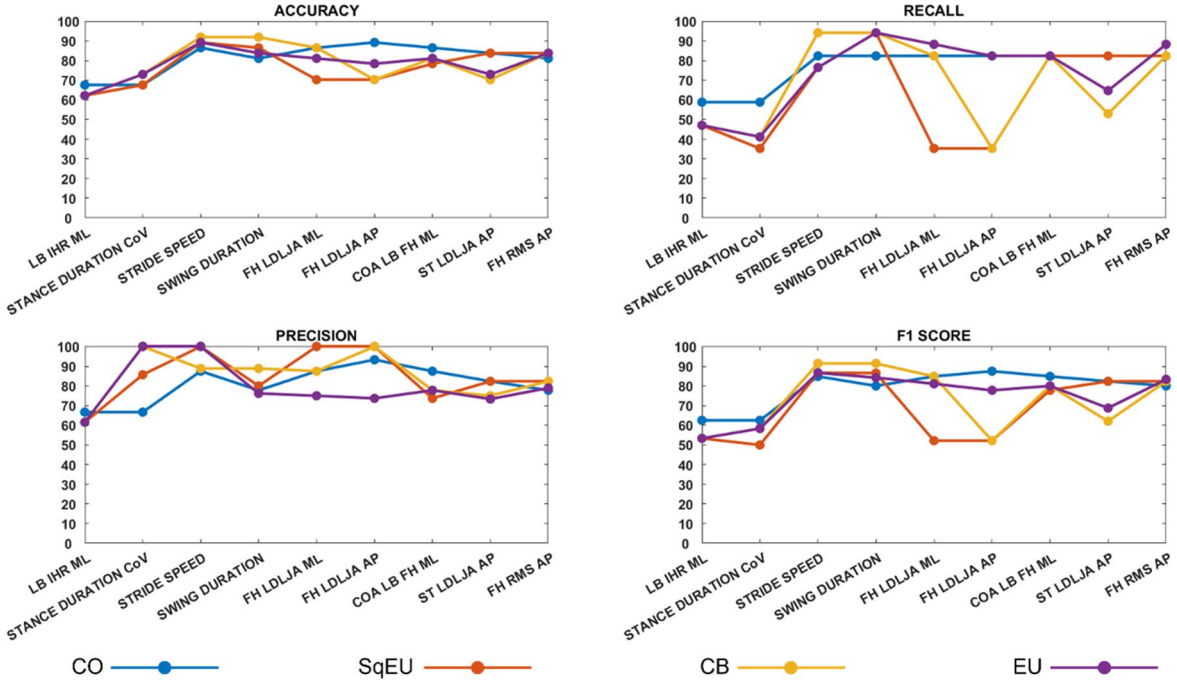

Fig. 5. Results of the unsupervised cluster on the Unsupervised Test Group. Accuracy, Recall, Precision, and F1-score are reported for the k-medoids using four distance metrics (CO, SqEU, CB, and EU) while adding one-by-one the feature ordered according to the total number of occurrences at the end of the SBS.

Distance metrics	Accuracy		Recall		Precision		F1-Score	
	ten runs	*Final test*	*ten runs*	*Final test*	*ten runs*	*Final test*	*ten runs*	*Final test*
CO	84±4%	81%	83±5.9%	82.3%	82.6±4%	77.7%	82.7±4.4%	80%
SqEU	75.8±8%	83.8%	64±19.2%	82.3%	82.5±12.6%	82.3%	70±12.3%	82.3%
CB	75.8±8.1%	83.8%	61±19.7%	82.3%	85.8±14%	82.3%	68.8±13.2%	82.3%
EU	77.7±9.2%	83.8%	69.5±20.5%	88.2%	81.4±11.3%	79%	73.1±13.5%	83.3%

Table 2. Unsupervised clustering performance of k-medoids using the SBS-derived feature subset across the ten SBS test sets (mean ± standard deviation) and on the independent Unsupervised Test Group (Final test) for each distance metric.

necessarily improve monotonically as further features are added, suggesting that the most frequently selected descriptors capture the core information needed for unsupervised separation between healthy participants and stroke patients, while additional features may contribute limited gains or introduce redundancy.

Table 2 reports the performance of k-medoids clustering using the SBS-derived feature subset when varying the distance metric, both across the ten runs (mean ± standard deviation computed over the ten test sets) and on the independent Unsupervised Test Group ("Final test"). Across the ten runs, cosine distance (CO) yielded the most favorable overall results, with the highest mean accuracy and F1-score and comparatively limited dispersion, whereas the remaining distances exhibited lower average accuracy, largely driven by less stable recall (with markedly larger run-to-run variability). When the same clustering approach was applied to the Unsupervised Test Group, performance remained consistently high across all distances, with comparable accuracy values and good recall/F1-score, indicating that the selected feature set preserves discriminative structure also in unseen participants and not only within the supervised SBS framework.

Discussion

The purpose of this study was to identify an optimal set of MIMU-based features able to distinguish between healthy participants and patients with stroke. This approach aims to characterize the gait patterns of people after a stroke event by identifying the features that most effectively capture deviations from physiological conditions. The methodological approach includes several strategies to overcome some of the limitations highlighted in a recent review on the topic and to improve the robustness and generalizability of the results[11]. First, a larger sample size was used to limit classification performance overestimation and improve results generalizability[39–41]. Wearable sensors, specifically MIMUs, were used to propose a model suitable for clinical practice. An instrumental setup was selected that balances the minimum number of required devices with the number of measurable useful features. Only clinically meaningful features were extracted, and a feature-selection approach was used to reduce the feature space while preserving clinical interpretability of the starting dataset. The dataset was split into training, validation, and test sets to perform algorithm tuning, feature selection, and to test their discriminative ability. The entire procedure was repeated while randomly changing the composition of the datasets to assess robustness across data splits. Different classifiers were used to identify features consistently retained across algorithms and those that were not. Finally, an unsupervised clustering technique was employed to verify whether the selected features retain discriminative structure in an unsupervised setting. Results suggest that the method proposed here can be used to select features that highlight differences between healthy and pathological locomotion (when using both supervised classification and unsupervised clustering), thereby highlighting the gait domains that most clearly differentiate stroke from healthy gait.

The set of 79 features analyzed in the present study has been selected based on the scientific literature on the topic[8,42,43]. Specifically, features that described spatiotemporal, symmetry, variability, stability, and smoothness domains of gait were considered. Features were chosen to characterize both general aspects of gait (i.e., those derived from shanks, and the lower back, LB, MIMUs) and the quality of movement of the upper body (i.e., those derived from the sternum, ST, and the forehead, FH, MIMUs). Feature selection procedures involved both statistical and machine learning approaches. The statistical approach reduced the initial number of features from 79 to 20, which were then further used for the machine learning feature selection approach, based on three different classification algorithms: KNN, SVM, and TREE. When using KNN, SVM, and TREE, different results were obtained in terms of both classification performance and of retained features after feature selection.

Specifically, when looking at the classification accuracy, SVM performed better in both the training and the test sets. The same result was found by Trabassi and colleagues[37] when classifying patients with Parkinson's disease and healthy people using a similar approach with respect to this study. These results suggest the potential effectiveness of SVM in detecting gait deviations in neurological patients. Concerning the other performance indexes obtained from the test sets, SVM showed the highest values except for classification precision, which was higher when using the KNN. The classification performance achieved in this study is slightly lower compared to those obtained in previous studies on patients with stroke[13–16,20,21] but there are several methodological differences to consider. First, previous studies often had significantly smaller sample sizes (from a minimum of 15 to a maximum of 58 participants), which can lead to overestimation of classification performance due to overfitting and random effects[44]. Moreover, the considered studies used multiple observations per participant rather than using a single representative data point for each participant (like the median value over several gait cycles and trials in this study). When using multiple data of the same participant, some may appear in both training and testing sets, thus increasing the risk of overestimating classification performance[45]. In some cases,

this issue was considered and avoided[15,16]. Differences in classification performance can also be attributed to the use of different features related to various gait domains, such as joint kinematics[16], as well as demographic differences among study groups[16]. Additionally, variations in data processing, such as focusing solely on the affected side of pathological participants[15], may have led to different results.

When examining the features selected across the ten runs, the three classifiers showed different selection patterns. In general, KNN retained the smallest number of features, SVM showed intermediate behavior, and TREE retained the largest number (Fig. 3). A similar algorithm-dependent variability has been reported in the literature: for example, Altilio and colleagues[24] tested multiple classifiers and quantified feature relevance by counting selection occurrences, showing that the retained subset can change depending on the model. This behavior is expected because feature selection is optimized with respect to each classifier's learning mechanism and inductive bias. KNN, being distance-based, is typically more sensitive to redundant or noisy predictors that can blur neighborhood structure and therefore tends to favor more compact subsets, whereas decision trees can exploit a larger pool of variables through hierarchical splits and interactions, often retaining more features; SVM commonly falls in between due to margin maximization and regularization. For this reason, although many studies adopt a single classifier, we used multiple algorithms to reduce the risk of deriving an algorithm-specific subset and to identify features that are consistently retained across learning paradigms.

This study highlights key methodological challenges when using machine learning for gait classification. Notably, classification performance varies when the same algorithm is applied to different datasets (Table 1; Fig. 2). This variation is particularly significant, given that the ten splits were not independent samples and may share similar participants, yet performance still varied, demonstrating that even minor changes in data can lead to different results. Therefore, the issue of limited representativeness due to small sample sizes in previous studies becomes pertinent. In contrast, this study enrolled more participants than those in earlier research, aligning with the recommendations by Jiao and colleagues[11]. Second, the optimal set of features to be used to discriminate between two different populations also depends on the algorithm used. Consequently, when using a single algorithm, the selected features are those that maximize classification performance for that specific algorithm in a defined dataset rather than a set of features capable of discriminating between two populations, which is often the actual objective.

The iterative feature selection technique used in this study, combined with the application of different machine learning algorithms tuned with a 5-fold cross-validation, enhances the generalizability of the results. This approach aims to improve reliability and reduce split-dependent effects, while optimizing classification performance.

A k-medoids clustering algorithm was chosen using four different distance metrics, with a deterministic medoids initialization algorithm to ensure repeatability of the results. The selected features were ranked by their number of occurrences and incrementally added to evaluate the incremental contribution of features to clustering performance as well as the performance of all the selected features. When looking at the clustering output using the whole set of features on the ten test sets, results are promising, lying in the range identified in the review by Jiao and colleagues[11] when using different supervised algorithms (i.e., 80–100%). Among the distance metrics used, the cosine distance achieved the best performance for all the classification performance indexes with the only exception of precision (see Table 2). However, in this context, precision was calculated as the ability to correctly classify healthy participants and thus may be considered less clinically relevant. The higher performance observed with the cosine distance, compared with the other metrics, likely reflects the different information emphasized by each distance measure. Indeed, cosine distance evaluates the orientation of n-dimensional vectors formed by the selected features measuring the angle between them, whereas the other metrics focus, albeit slightly differently, on the absolute distance between these vectors. In the first case, the magnitude of the vectors does not influence the results whereas in the other metrics magnitude contributes to the distance. In our dataset, cosine distance seems to capture group differences more consistently. Moreover, the cosine distance exhibited the smallest standard deviation across all performance indexes over the ten test sets, suggesting greater consistency across runs in this dataset.

On the test sets, performance peaked when the model relied on only three features: the improved harmonic ratio in the medio-lateral direction, the coefficient of variation of the stance phase, and stride speed. Thus, a further reduction of the number of features used to discriminate between healthy participants and stroke patients may improve classification performance. When examining the results obtained on the Unsupervised Test Group using all features and the cosine distance, the outcomes appear consistent with those from the ten runs. Conversely, results obtained with other distance metrics show higher performance metrics. Nevertheless, this is not surprising given the large standard deviation observed over the ten runs with these metrics, indicating a high variability. It is likely that in some of the ten test sets, the clustering performance aligned with those from the Unsupervised Test Group.

Notably, the highest accuracy, precision, and F1-score when using the cosine distance were obtained using six features. Those included the duration of the stance phase and the movement smoothness of the head in the anterior-posterior and medio-lateral directions. However, the differences in results between using three versus six features were modest (+3%, +6%, and +3% for accuracy, precision, and F1-score, respectively). Recall values, in contrast, remained unchanged after the third feature was added. These findings suggest that even with an unsupervised clustering technique, variability across splits can be partly mitigated. The results presented here may be summarized as follows:

- Among all 79 features considered in this work, nine features (see Fig. 3) seem to be sufficient to discriminate between healthy participants and participants with stroke with fair-to-good classification performance.
- When using an unsupervised clustering technique based on the distances between data points, the cosine distance showed the most consistent performance across runs.

- Generally, to maximize the classification results, three features are sufficient; nevertheless, other features may carry discriminative information and should be considered.

The procedure applied in this study has yielded some insightsinto the most discriminative gait-related features that differentiate between healthy individuals and people with stroke. Importantly, these features are not merely selected to optimize classification performance but reflect specific physiological and pathological mechanisms underlying post-stroke gait dysfunction. In particular, the most frequently selected features belong to the spatiotemporal, symmetry, stability, and smoothness domains, which are known to be directly affected by stroke-related impairments such as hemiparesis, altered neuromuscular control, and impaired postural regulation during walking[42,43,46–50]. Alterations in spatiotemporal gait parameters, including reduced stride speed and increased variability of stance-phase duration, represent hallmark manifestations of post-stroke gait. Increased temporal variability reflects reduced gait automaticity and impaired inter-limb coordination, which have been associated with poorer functional outcomes and increased fall risk in people with stroke. From a clinical perspective, these features capture deficits in motor control rather than compensatory strategies alone, making them particularly relevant for pathological analysis and neurorehabilitation monitoring.

Notably, none of the selected features represented asymmetry in spatiotemporal parameters. While gait asymmetry is a widely recognized alteration after stroke, it is often characterized by varying patterns[42,43,46]. These diverse trends in spatiotemporal feature asymmetry may have reduced their discriminative value. However, an asymmetry feature based on the frequency content of the acceleration measured at the lower back in the mediolateral direction was still included in the selected features. Generally, asymmetry at the trunk level has been widely analyzed in stroke patients and has been reported to significantly differ compared to healthy[42,43] controls. This is true not only when using a single MIMU at the lumbar level, but as well as multiple sensors are used and when different aspects of trunk movement are analyzed[43]. Accordingly, some of the features considered in the present work capture aspects of trunk movement stability and symmetry.

Finally, trunk smoothness (at different levels and in different directions) emerged as a discriminative gait domain. Reduced movement smoothness reflects impaired sensorimotor integration, diminished anticipatory postural adjustments, and inefficient balance control during locomotion. To the authors' knowledge, only two prior studies have directly measured trunk smoothness in stroke patients, albeit using different[51] equipment or metrics[52]. Notably, trunk and, in particular, head movement smoothness, were identified as key discriminative features. While traditionally less emphasized in post-stroke gait analysis, head movement smoothness is closely linked to the control of dynamic gaze stabilization and vestibular–spinal integration during walking[53–56]. Its alteration may therefore indicate deficits extending beyond lower-limb motor impairment, involving higher-level balance and sensory reweighting mechanisms. This finding opens new avenues for incorporating head movement analysis into routine post-stroke gait assessments, and could represent a promising target for future rehabilitation strategies to improve dynamic gaze and postural stability.

Conclusion

The findings should be interpreted in light of several limitations. First, we focused on three conventional classifiers (KNN, SVM, and decision trees); other modelling approaches (e.g., ensemble methods or deep-learning models) could be explored in future work to verify whether similar feature subsets emerge. Second, walking speed was included among the candidate features. Because speed influences many gait descriptors, part of the discriminative performance may reflect speed-related differences between groups. Future studies should therefore replicate the analysis under speed-matched conditions or by explicitly controlling for speed (e.g., stratification or covariate adjustment), although this may reduce representativeness or sample size in heterogeneous cohorts[37]. Third, the correlation-based redundancy filtering adopted prior to feature selection may have excluded potentially informative features, depending on the chosen threshold and arbitration rules. Nevertheless, the final subset still supported consistent discrimination between healthy controls and participants with stroke in both supervised and unsupervised analyses, indicating practical usefulness despite possible feature omissions. Finally, the stroke cohort included acute, subacute, and chronic participants; while this heterogeneity may have increased variability and reduced peak performance, it also suggests that the identified features capture clinically meaningful aspects of gait across different stages after stroke.

In conclusion, the present study demonstrates the potential of machine learning in identifying key features of post-stroke gait dysfunctions. The highlighted features—spatiotemporal features, gait variability and trunk movement symmetry, stability, and smoothness—not only enhance our understanding of post-stroke gait dysfunctions but also provide practical markers for clinical assessment and rehabilitation. It would be beneficial for future work to focus on refining machine learning models to support real-time gait analysis and expanding their application in diverse clinical settings, ensuring their integration into personalized and effective rehabilitation strategies for patients with stroke and neurological conditions.

Data availability

The data associated with this paper are not publicly available but are available from the corresponding author on reasonable request.

Received: 28 August 2025; Accepted: 5 March 2026
Published online: 09 March 2026

References

1. Sacco, R. L. et al. An updated definition of stroke for the 21st century. *Stroke* **44**, 2064–2089 (2013).

2. Martin, S. S. et al. Heart Disease and Stroke Statistics: A Report of US and Global Data From the American Heart Association. *Circulation* (2024). https://doi.org/10.1161/CIR.0000000000001209 (2024) doi:10.1161/CIR.0000000000001209.
3. Wang, W. et al. Prevalence, incidence, and mortality of stroke in China. *Circulation* https://doi.org/10.1161/CIRCULATIONAHA.116.025250 (2017).
4. Kim, Y. W. Update on stroke rehabilitation in motor impairment. *Brain Neurorehabil.* **15**, e12 (2022).
5. Selves, C., Stoquart, G. & Lejeune, T. Gait rehabilitation after stroke: Review of the evidence of predictors, clinical outcomes and timing for interventions. *Acta Neurol. Belg.* **120**, 783790 (2020).
6. Kinoshita, S., Abo, M., Okamoto, T. & Tanaka, N. Utility of the revised version of the ability for basic movement scale in predicting ambulation during rehabilitation in poststroke patients. *J. Stroke Cerebrovasc. Dis.* **26**, 1663–1669 (2017).
7. Hutabarat, Y., Owaki, D. & Hayashibe, M. Recent advances in quantitative gait analysis using wearable sensors: A review. *IEEE Sens. J.* **21**, 26470–26487 (2021).
8. Mohan, D. M. et al. Assessment Methods of Post-stroke Gait: A Scoping Review of Technology-Driven Approaches to Gait Characterization and Analysis. *Frontiers Neurology* **12**, 650024 (2021).
9. Kim, G. J., Parnandi, A., Eva, S. & Schambra, H. The use of wearable sensors to assess and treat the upper extremity after stroke: A scoping review. *Disabil. Rehabil.* **44**, 6119–6138 (2022).
10. Picerno, P. et al. Wearable inertial sensors for human movement analysis: a five-year update. *Expert Rev. Med. Devices.* **18**, 79–94 (2021).
11. Jiao, Y., Hart, R., Reading, S. & Zhang, Y. Systematic review of automatic post-stroke gait classification systems. *Gait Posture.* **109**, 259–270 (2024).
12. Boukhennoufa, I., Zhai, X., Utti, V., Jackson, J. & McDonald-Maier, K. D. Wearable sensors and machine learning in post-stroke rehabilitation assessment: A systematic review. *Biomed. Signal Process. Control.* **71**, 103197 (2022).
13. Lee, J., Park, S. & Shin, H. Detection of hemiplegic walking using a wearable inertia sensing device. *Sens. (Basel)* **18**, 1736 (2018).
14. Hsu, W.-C. et al. Can trunk acceleration differentiate stroke patient gait patterns using time- and frequency-domain features?. *Applied Sciences* **11**, 1541 (2021).
15. Mannini, A., Trojaniello, D., Cereatti, A. & Sabatini, A. M. A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and Huntington's disease patients. *Sensors (Basel)* **16**, 134 (2016).
16. Scheffer, C. & Cloete, T. Inertial motion capture in conjunction with an artificial neural network can differentiate the gait patterns of hemiparetic stroke patients compared with able-bodied counterparts. *Comput. Methods Biomech. Biomed. Engin.* **15**, 285–294 (2012).
17. Wang, L., Sun, Y., Li, Q., Liu, T. & Yi, J. Two shank-mounted IMUs-based gait analysis and classification for neurological disease patients. *IEEE Rob. Autom. Lett.* **5**, 1970–1976 (2020).
18. Altilio, R., Paoloni, M. & Panella, M. Selection of clinical features for pattern recognition applied to gait analysis. *Med. Biol. Eng. Comput.* **55**, 685–695 (2017).
19. Iosa, M., Picerno, P., Paolucci, S. & Morone, G. Wearable inertial sensors for human movement analysis. *Expert Rev. Med. Devices.* **13**, 641–659 (2016).
20. Wang, F.-C. et al. Detection and classification of stroke gaits by deep neural networks employing inertial measurement units. *Sensors (Basel)* **21**, 1864 (2021).
21. Mathur, D. & Bhatia, D. Gait classification of stroke survivors - An analytical study. *J. Interdiscip. Math.* **25**, 163–181 (2022).
22. Sung, J. et al. Classification of stroke severity using clinically relevant symmetric gait features based on recursive feature elimination with cross-validation. *IEEE Access* **10**, 119437–119447 (2022).
23. Altilio, R., Liparulo, L., Proietti, A., Paoloni, M. & Panella, M. A genetic algorithm for feature selection in gait analysis. in *IEEE Congress on Evolutionary Computation (CEC)* 4584–4591 (2016). 4584–4591 (2016). (2016). https://doi.org/10.1109/CEC.2016.7744374
24. Altilio, R., Rossetti, A., Fang, Q., Gu, X. & Panella, M. A comparison of machine learning classifiers for smartphone-based gait analysis. *Med. Biol. Eng. Comput.* **59**, 535–546 (2021).
25. Iosa, M. et al. Artificial neural network analyzing wearable device gait data for identifying patients with stroke unable to return to work. *Front. Neurol.* **12**, 650542 (2021).
26. Holden, M. K., Gill, K. M., Magliozzi, M. R., Nathan, J. & Piehl-Baker, L. Clinical gait assessment in the neurologically impaired. Reliability and meaningfulness. *Phys. Ther.* **64**, 35–40 (1984).
27. Folstein, M. F., Folstein, S. E. & McHugh, P. R. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).
28. Bergamini, E. et al. Estimating orientation using magnetic and inertial sensors and different sensor fusion approaches: accuracy assessment in manual and locomotion tasks. *Sens. (Basel).* **14**, 18625–18649 (2014).
29. Kavanagh, J. J. & Menz, H. B. Accelerometry: A technique for quantifying movement patterns during walking. *Gait Posture* **28**, 1–15 (2008).
30. Madgwick, S. O. H., Harrison, A. J. L. & Vaidyanathan, A. Estimation of IMU and MARG orientation using a gradient descent algorithm. *IEEE Int Conf Rehabil Robot* 5975346 (2011). (2011).
31. Bertoli, M. et al. Estimation of spatio-temporal parameters of gait from magneto-inertial measurement units: Multicenter validation among Parkinson, mildly cognitively impaired and healthy older adults. *BioMed. Eng. OnLine* **17**, 58 (2018).
32. Menz, H. B., Lord, S. R. & Fitzpatrick, R. C. Acceleration patterns of the head and pelvis when walking on level and irregular surfaces. *Gait Posture* **18**, 35–46 (2003).
33. Buckley, C., Galna, B., Rochester, L. Mazzà, C. Attenuation of upper body accelerations during gait: Piloting an innovative assessment tool for Parkinson's Disease. *BioMed. Res. Int.* **2015**, 865873 (2015).
34. Pasciuto, I., Bergamini, E., Iosa, M., Vannozzi, G. & Cappozzo, A. Overcoming the limitations of the harmonic ratio for the reliable assessment of gait symmetry. *J. Biomech.* **53**, 84–89 (2017).
35. Melendez-Calderon, A., Shirota, C. & Balasubramanian, S. Estimating movement smoothness from inertial measurement units. *Front. Bioeng. Biotechnol.* **8**, 558771 (2020).
36. Zifchock, R. A., Davis, I., Higginson, J. & Royer, T. The symmetry angle: A novel, robust method of quantifying asymmetry. *Gait Posture* **27**, 622–627 (2008).
37. Trabassi, D. et al. Machine learning approach to support the detection of Parkinson's Disease in IMU-based gait analysis. *Sensors* **22**, 3700 (2022).
38. Pavan, K. K., Rao, A. A., Rao, A. V. D. & Sridhar, G. R. Single pass seed selection algorithm for k-Means. *JCS* **6**, 60–66 (2010).
39. McCrum, C., van Beek, J., Schumacher, C., Janssen, S. & Van Hooren, B. Sample size justifications in gait & posture. *Gait & Posture* **92**, 333–337 (2022).
40. Lakens, D. Sample size justification. *Collabra: Psychology* **8**, 33267 (2022).
41. Trabassi, D. et al. Optimizing rare disease gait classification through data balancing and generative AI: Insights from Hereditary Cerebellar Ataxia. *Sensors* **24**, 3613 (2024).
42. Tramontano, M. et al. Dynamic stability, symmetry, and smoothness of gait in people with neurological health conditions. *Sensors* **24**, 2451 (2024).
43. Bergamini, E. et al. Multi-sensor assessment of dynamic balance during gait in patients with subacute stroke. *J Biomech* **61**, 208–215 (2017).

44. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLOS ONE* **14**, e0224365 (2019).
45. Chaibub Neto, E. et al. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *npj Digit. Med.* **2**, 1–6 (2019).
46. Little, V. L., Perry, L. A., Mercado, M. W., Kautz, S. A. & Patten, C. Gait asymmetry pattern following stroke determines acute response to locomotor task. *Gait Posture.* **77**, 300–307 (2020).
47. Patterson, K. K., Gage, W. H., Brooks, D., Black, S. E. & McIlroy, W. E. Evaluation of gait symmetry after stroke: A comparison of current methods and recommendations for standardization. *Gait Posture.* **31**, 241–246 (2010).
48. Balasubramanian, C. K., Neptune, R. R. & Kautz, S. A. Variability in spatiotemporal step characteristics and its relationship to walking performance post-stroke. *Gait Posture.* **29**, 408–414 (2009).
49. Kim, C. M. & Eng, J. J. Symmetry in vertical ground reaction force is accompanied by symmetry in temporal but not distance variables of gait in persons with stroke. *Gait Posture.* **18**, 23–28 (2003).
50. Bowden, M. G., Balasubramanian, C. K., Behrman, A. L. & Kautz, S. A. Validation of a speed-based classification system using quantitative measures of walking performance poststroke. *Neurorehabil. Neural Repair.* **22**, 672–675 (2008).
51. Germanotta, M., Iacovelli, C. & Aprile, I. Evaluation of gait smoothness in patients with stroke undergoing rehabilitation: Comparison between two metrics. *Int. J. Environ. Res. Public Health.* **19**, 13440 (2022).
52. Garcia, F. et al. Movement smoothness in chronic post-stroke individuals walking in an outdoor environment—A cross-sectional study using IMU sensors. *PLoS One.* **16**, e0250100 (2021).
53. Bent, L. R., Inglis, J. T. & McFadyen, B. J. Vestibular contributions across the execution of a voluntary forward step. *Exp. Brain Res.* **143**, 100–105 (2002).
54. Bent, L. R., McFadyen, B. J. & Inglis, J. T. Vestibular contributions during human locomotor tasks.. *Exerc. Sport Sci. Rev.* **33**, 107 (2005).
55. Vestibulospinal and Reticulospinal Neuronal Activity During Locomotion in the Intact Cat. I. Walking on a Level Surface | Journal of Neurophysiology | American Physiological Society. https://journals.physiology.org/doi/full/10.1152/jn.2000.84.5.2237?rfr_dat=cr_pub++0pubmed&url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org.
56. Frontiers | Insufficiencies in sensory systems reweighting is associated with walking impairment severity in chronic stroke: an observational cohort study. (2023). https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.1244657/full.

Author contributions

Brasiliano Paolo: Conceptualization, Methodology, Software, Validation, Formal Analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. Orejel-Bustos Amaranta: Investigation, Data Curation. Belluscio Valeria: Investigation, Writing - Review & Editing. Cereatti Andrea: Software Della Croce Ugo: Software Trabassi Dante: Methodology, Writing - Review & Editing. Salis Francesca: Software Tramontano Marco: Resources, Writing - Review & Editing, Funding Acquisition. Buzzi Maria Gabriella: Resources, Vannozzi Giuseppe: Writing - Review & Editing Bergamini Elena: Supervision, Project administration, Writing - Review & Editing, Funding Acquisition.

Funding

This study was supported by the Italian Ministry of Health (GR-2019-12370757).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-43666-7.

Correspondence and requests for materials should be addressed to P.B.