Gait & Posture 111 (2024) 105–121

Contents lists available at ScienceDirect

Gait & Posture

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

Machine learning applied to gait analysis data in cerebral palsy and stroke: A systematic review

Farshad Samadi Kohnehshahri [a,b], Andrea Merlo [b,*], Davide Mazzoli [b], Maria Chiara Bò [b,c], Rita Stagni [a]

[a] Department of Electronic and Information Engineering, University of Bologna, Italy
[b] Gait and Motion Analysis Laboratory, Sol et Salus Hospital, Torre Pedrera, Rimini, Italy
[c] Merlo Bioengineering, Parma, Italy

ARTICLE INFO

Keywords:
Machine learning
Gait analysis
Cerebral palsy
Stroke
Systematic review

ABSTRACT

Background: Among neurological pathologies, cerebral palsy and stroke are the main contributors to walking disorders. Machine learning methods have been proposed in the recent literature to analyze gait data from these patients. However, machine learning methods still fail to translate effectively into clinical applications. This systematic review addressed the gaps hindering the use of machine learning data analysis in the clinical assessment of cerebral palsy and stroke patients.

Research Question: What are the main challenges in transferring proposed machine learning methods to clinical applications?

Methods: PubMed, Web of Science, Scopus, and IEEE databases were searched for relevant publications on machine learning methods applied to gait analysis data from stroke and cerebral palsy patients until February the 23rd, 2023. Information related to the suitability, feasibility, and reliability of the proposed methods for their effective translation to clinical use was extracted, and quality was assessed based on a set of predefined questions.

Results: From 4120 resulting references, 63 met the inclusion criteria. Thirty-one studies used supervised, and 32 used unsupervised machine learning methods. Artificial neural networks and k-means clustering were the most used methods in each category. The lack of rationale for features and algorithm selection, the use of unrepresentative datasets, and the lack of clinical interpretability of the clustering outputs were the main factors hindering the clinical reliability and applicability of these methods.

Significance: The literature offers numerous machine learning methods for clustering gait data from cerebral palsy and stroke patients. However, the clinical significance of the proposed methods is still lacking, limiting their translation to real-world applications. The design of future studies must take into account clinical question, dataset significance, feature and model selection, and interpretability of the results, given their criticality for clinical translation.

1. Introduction

Neurological conditions account for two-thirds of mobility limitations in individuals with walking disorders [1]. Among the neurological pathologies, Cerebral Palsy (CP), with a prevalence of 1.5 to 3 cases per 1000 live births in Europe, is recognized as the predominant cause of movement disorders in children [2], while stroke stands as the primary contributor to disability among older adults, estimated to reach from 1.1 in 2000 to 1.5 million per year in 2025 [3]. Due to the prevalence of these pathologies, most literature on gait analysis (GA) clinical applications [4,5] refers to CP and stroke, comprising more than 17% of total publications in clinical GA over the past decade, as per PubMed search results.

In assessing walking impairments of CP and stroke patients, GA holds significant clinical relevance, providing a quantitative description of gait functional alterations and retaining the potential to support clinical

* Correspondence to: Gait and Motion Analysis Laboratory OPA Sol et Salus, Viale San Salvador, 204, Rimini, Torre Pedrera 47922, Italy.
E-mail addresses: Farshad.samadi2@unibo.it (F. Samadi Kohnehshahri), a.merlo@soletsalus.com (A. Merlo), d.mazzoli@soletsalus.com (D. Mazzoli), chiarabo88@gmail.com (M.C. Bò), rita.stagni@unibo.it (R. Stagni).

https://doi.org/10.1016/j.gaitpost.2024.04.007
Received 16 January 2024; Received in revised form 8 March 2024; Accepted 8 April 2024
Available online 10 April 2024
0966-6362/

assessment effectively [6]. Functional diagnosis, disease progression monitoring, treatment plan, and pre- and post-operative evaluation are among the current applications of GA in clinical settings [5,7–9]. Mainly, GA-based support to clinical decision-making can be of utmost relevance when non-conservative treatments are designed, such as selective neurotomies and neuro-orthopedic surgery in these pathologies [10,11].

The huge volume and diversity of GA data, including kinematics, spatio-temporal parameters, dynamics, and muscle activation patterns, poses significant challenges for clinicians in terms of time required for interpretation and consensus in clinical decision-making [12,13]. To overcome these challenges, data analysis techniques have been developed to analyse [14,15], summarize [16–18], and cluster [19,20] GA data. Beyond the efficacy of the traditional data analysis techniques in clinical GA, the difficulties in capturing complex non-linear relationships within multi-dimensional and highly variable gait data, reducing the amount of data and the time-consuming process of manual key features selection, which is also prone to biases, suggests the necessity for more effective approaches [21,22]. The ability of machine learning (ML) algorithms to capture nonlinear patterns, handle vast amounts of data, engineering features, and generalize makes them potentially beneficial in the analysis of gait data [23]. Supervised and unsupervised methods are two primary classes of ML algorithms. Supervised methods involve training a model on labeled data, where the algorithm learns to predict outcomes based on input features and corresponding target variables. In contrast, unsupervised methods deal with unlabeled data, where the algorithm seeks to uncover hidden patterns or structures within the data without explicit guidance on the output.

In the literature on the use of GA data acquired from individuals with CP and stroke, ML algorithms have been introduced for a wide range of clinical applications, such as distinguishing between healthy and pathological gait [24–27], discriminating between different pathologies [28–31], identifying distinct patterns within the same disorder [32–36], assessing the effectiveness of interventions [37–39], and recommending appropriate treatment strategies [40]. However, the translation of ML methods into clinical applications poses three significant challenges: i) *suitability*, as the appropriateness of the algorithm to meet clinical application requirements and the ability to extract clinically relevant and usable information; ii) *feasibility*, as the practicality of these methods in real-world clinical settings, considering available resources and technical supports; iii) *reliability*, as the consistency and the robustness of ML algorithms when applied to the specific gait data [41]. These challenges must be properly addressed to support the application of the proposed models in the clinical setting.

To analyse this issue, we designed and implemented a systematic review of the available literature on ML methods applied to GA data from CP and stroke individuals. We focused on samples and dataset quality, feature engineering, method selection, algorithm design, validation, and clinical relevance of findings, and identified the strengths and limitations of the study designs. This will provide a reference framework for researchers in the design of future studies on the use of ML techniques on GA data when targeting clinical applications.

2. Methods

The present systematic review follows the guidelines of the PRISMA statement [42].

2.1. Study selection and research criteria

The search was performed in 4 databases (i.e., PubMed, Web of Science, Scopus, and IEEE) and was completed on February the 23rd, 2023. The search string was adapted to align with each database's specific search protocols and syntax requirements (Table 1). The search results from the different databases were combined into a single list after removing duplicate articles. In addition to the systematic searching,

Table 1
Search Strings per database.

Database	Research String
PubMed	("cerebral palsy"[MeSH Terms] OR ("cerebral"[All Fields] AND "palsy"[All Fields]) OR "cerebral palsy"[All Fields] OR ("stroke"[MeSH Terms] OR "stroke"[All Fields] OR "strokes"[All Fields] OR "stroke s"[All Fields]) OR "poststroke"[All Fields]) AND ("gait"[MeSH Terms] OR "gait"[All Fields] OR (("lower extremity"[MeSH Terms] OR ("lower"[All Fields] AND "extremity"[All Fields]) OR "lower extremity"[All Fields] OR ("lower"[All Fields] AND "limb"[All Fields]) OR "lower limb"[All Fields]) AND ("muscle s"[All Fields] OR "muscles"[MeSH Terms] OR "muscles"[All Fields] OR "muscle"[All Fields]))) AND ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields] OR "data-driven"[All Fields] OR ("deep learning"[MeSH Terms] OR ("deep"[All Fields] AND "learning"[All Fields]) OR "deep learning"[All Fields]) OR ("classification"[MeSH Terms] OR "classification"[All Fields] OR "classifications"[All Fields] OR "classification"[MeSH Subheading] OR "classification s"[All Fields] OR "classificator"[All Fields] OR "classificators"[All Fields]) OR ("classifiable"[All Fields] OR "classification"[MeSH Terms] OR "classification"[All Fields] OR "classified"[All Fields] OR "classify"[All Fields] OR "classifying"[All Fields] OR "classifier"[All Fields] OR "classifier s"[All Fields] OR "classifiers"[All Fields] OR "classifies"[All Fields]) OR ("cluster analysis"[MeSH Terms] OR ("cluster"[All Fields] AND "analysis"[All Fields]) OR "cluster analysis"[All Fields]) OR ("cluster analysis"[MeSH Terms] OR ("cluster"[All Fields] AND "analysis"[All Fields]) OR "cluster analysis"[All Fields] OR "clustering"[All Fields] OR "clusterings"[All Fields] OR "cluster"[All Fields] OR "cluster s"[All Fields] OR "clustered"[All Fields] OR "clusterization"[All Fields] OR "clusters"[All Fields]) OR ("proc iapr int conf pattern recogn"[Journal] OR "pattern recognit"[Journal] OR "pattern recognit dagm"[Journal] OR "pattern recognit 2021"[Journal] OR ("pattern"[All Fields] AND "recognition"[All Fields]) OR "pattern recognition"[All Fields]) OR ("categoric"[All Fields] OR "categorical"[All Fields] OR "categorically"[All Fields] OR "categorisation"[All Fields] OR "categorisations"[All Fields] OR "categorise"[All Fields] OR "categorised"[All Fields] OR "categorises"[All Fields] OR "categorising"[All Fields] OR "categorizations"[All Fields] OR "categorize"[All Fields] OR "categorized"[All Fields] OR "categorizer"[All Fields] OR "categorizers"[All Fields] OR "categorizes"[All Fields] OR "categorizing"[All Fields] OR "classification"[MeSH Terms] OR "classification"[All Fields] OR "categorization"[All Fields]) OR (("feature s"[All Fields] OR "featured"[All Fields] OR "features"[All Fields] OR "featuring"[All Fields] OR "protein domains"[MeSH Terms] OR ("protein"[All Fields] AND "domains"[All Fields]) OR "protein domains"[All Fields] OR "feature"[All Fields]) AND ("extract"[All Fields] OR "extract s"[All Fields] OR "extractabilities"[All Fields] OR "extractability"[All Fields] OR "extractable"[All Fields] OR "extractables"[All Fields] OR "extractant"[All Fields] OR "extractants"[All Fields] OR "extracted"[All Fields] OR "extractibility"[All Fields] OR "extractible"[All Fields] OR "extracting"[All Fields] OR "extraction"[All Fields] OR "extractions"[All Fields] OR "extractive"[All Fields] OR "extractives"[All Fields] OR "extracts"[All Fields])))
Web of Science	((ALL=("cerebral palsy" OR Stroke OR poststroke)) AND ALL=(Gait OR "Lower limb muscle$")) AND ALL=("machine learning" OR "data driven" OR "deep learning" OR Classification OR Classifying OR "cluster analysis" OR clustering OR "pattern recognition" OR categorization OR "feature extraction")
Scopus	TITLE-ABS-KEY ("cerebral palsy" OR stroke OR poststroke) AND (gait OR "Lower limb muscle*") AND ("machine learning" OR "data-driven" OR "deep learning" OR classification OR classifying OR "cluster analysis" OR clustering OR "pattern recognition" OR categorization OR "feature extraction")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (PUBSTAGE, "final"))
IEEE Explore	(("All Metadata": "Cerebral palsy" OR "All Metadata": Stroke OR "All Metadata": Poststroke) AND ("All Metadata": Gait OR "All Metadata": "Lower limb muscle*") AND ("All Metadata":"machine learning" OR "All Metadata": "data-driven" OR "All Metadata": "deep learning" OR "All Metadata": Classification OR "All Metadata": Classifying OR "All

(continued on next page)

Table 1 (*continued*)

Database	Research String
	Metadata":"cluster analysis" OR "All Metadata": clustering OR "All Metadata":"pattern recognition" OR "All Metadata": categorization OR "All Metadata":"feature extraction"))

hand searching was also conducted using the selected databases and reference list of included studies to further enhance the comprehensiveness of the review.

Specific inclusion and exclusion criteria were applied to identify studies that are directly relevant to the research objectives of this review.

The inclusion criteria were:

1. Studies classifying gait impairment(s)/deviation(s) and describing or allocating gait variables into categories, groups, or clusters;
2. Studies using ML for processing data;
3. Studies involving human participants with CP or stroke;
4. Full papers;
5. Articles published in the English language.

The exclusion criteria were:

1. Studies analyzing upper limb motion;
2. Studies focusing on other neurological diseases (e.g., neurodegenerative diseases);
3. Studies involving the classification of walking-related conditions or events not directly related to clinical GA (e.g., actigraphy, risk of fall, identification of events, sport biomechanics, gait for biometrics, security, etc.);
4. Studies related to robotic-aided rehabilitation systems.

During study selection, titles and abstracts of all retrieved papers were screened independently by two reviewers. To decrease the risk of bias, the Rayyan software, a web-based tool specifically designed for this purpose, was used. Articles meeting the inclusion criteria were retrieved for full-text analysis.

2.2. Data extraction

One reviewer performed the data extraction using a structured table to extract all relevant features for the implementation of a clinically effective ML approach for analysing gait data: Reference; Pathologies; Main objective; Sample size; Measurement tool; Gait parameters; Analysed planes; ML Algorithms; Group numbers; Assessment methods.

2.3. Quality assessment

One reviewer assessed the quality of the included studies responding to a set of quality questions. Due to the inherent methodological differences in the implementation and the evaluation of supervised and unsupervised methods, a specific set of quality questions was defined for each of the 2 sub-groups (i.e., 16 questions for supervised, and 15 for unsupervised methods). Supervised methods questions were organized into seven categories: i) study design, ii) sample representativeness, iii) data acquisition, iv) features, v) dataset, vi) ML algorithm, vii) results validation and limitations. For unsupervised methods, quality questions

Table 2
Quality Assessment Questions sets for supervised and unsupervised methods.

Questionnaire for studies based on supervised methods:		Questionnaire for studies based on unsupervised methods:	
Study Design	QS1 Is the study design appropriate for the research question?	Study Design	QU1 Is the study design appropriate for the research question?
	QS2 Are the objectives and research questions clearly stated?		QU2 Are the objectives and research questions clearly stated?
Samples Representativeness	QS3 Is the representativeness of the sample justified and outlined in the study?	Samples Representativeness	QU3 Is the representativeness of the sample justified and outlined in the study?
Data Acquisition	QS4 Are the details of the data collection method clearly described?	Data Acquisition	QU4 Are the details of the data collection method clearly described?
Features	QS5 Are the selected gait features well-defined and clearly described?	Features	QU5 Are the selected gait features well-defined and clearly described?
	QS6 Are justification or rationale provided regarding the selection of features?		QU6 Are justification or rationale provided regarding the selection of features?
Dataset	QS7 Is there a clear description of data sets for the ML models?	ML Algorithm	QU7 Is there a justification for the selection of ML algorithms in the study?
	QS8 Are the models tested on external or independent datasets?		QU8 Are the details of the ML models clearly reported?
ML Algorithm	QS9 Is there a justification for the selection of ML algorithms in the study?	Results Validation, & Limitations & Clinical Relevance	QU9 Were the appropriate validation methods applied to evaluate the proposed model?
	QS10 Are the details of the ML models clearly reported?		QU10 Are the clinical explainability of the results addressed in the study?
Results Validation & Limitations	QS11 Were appropriate performance metrics used to evaluate the proposed model?		QU11 Are the conclusions clearly supported by the results?
	QS12 Are the conclusions clearly supported by the results?		QU12 Are the advantages and limitations of the chosen algorithms discussed?
	QS13 Are the advantages and limitations of the chosen algorithms discussed?		QU13 Are the results compared to the clinical benchmarks?
	QS14 Are the results compared to state-of-the-art approaches or benchmarks?		QU14 Are the limitations of the study clearly described?
	QS15 Are the limitations of the study clearly described?		QU15 Does the study add value to the state-of-the-art?
	QS16 Does the study add value to the state-of-the-art?		

were organized in six categories: i) study design, ii) samples representativeness, iii) data acquisition, iv) features, v) ML algorithm, vi) results validation, limitations and clinical relevance. The two sets of quality questions are reported in Table 2.

Each question was scored on a three-level basis: 1 for yes, 0.5 for limited details, and 0 for no [43,44]. The total score was computed per study to assess the overall quality and expressed as a percentage of the relative full score. Studies were classified into three categories: high quality (total score higher than 80%), medium quality (total score comprised between 51% and 79%), and low quality (total score lower than 50%) [43].

To offer a comprehensive overview of the challenges to the translation of ML methods into clinical use, three properties were specifically analysed: i) *suitability*, ii) *reliability*, iii) *feasibility*. To grade how these properties are addressed in the different studies, specific quality questions were associated with each property.

Questions and aspects associated to the suitability are:

- Feature selection (QS5, QS6, QU5, QU6)
- ML algorithm selection (QS9, QU7)
- Algorithms and Study limitations (QS13, QS15, QU12, QU14)
- Clinical explainability of the results (QU10)

Those linked to the feasibility:

- Sample representativeness (QS3, QU3)
- Data acquisition (QS4, QU4)
- ML algorithm selection (QS9, QU7)
- Algorithms and Study limitations (QS13, QS15, QU12, QU14)

And, for reliability:

- Sample representativeness (QS3, QU3)
- Data acquisition (QS4, QU4)
- Feature definition (QS5, QU5)
- Dataset (QS7, QS8)
- ML algorithm selection and details (QS9, QS10, QU7, QU8)
- Evaluation indices (QS11, QS12, QS14, QU9, QU11, QU13)
- Study limitations (QS15, QU14)

Based on the percentage of the studies that provided full answers to each question, we categorized the questions into three levels. Questions fully addressed in over 80% of the studies were assigned a high-quality rating, those answered between 51% and 79% were categorized as medium quality, and questions addressed in less than 50% of the studies were classified as low-quality.

For each property, we determined the ratio of questions at each level (low, medium, high) to the total number of questions linked to that property. These ratios are presented as percentages to evaluate the extent to which each property is addressed in studies. (For more details, please refer to the supplementary materials in Table S3).

3. Results

The initial search process yielded 4120 records through the defined search string and hand searching, which resulted in 2450 unique articles after removing duplicates. During the screening of titles and abstracts, conflicts were found in 45 papers. These conflicts were resolved through discussion, leading to a consensus for 40 articles, of which 20 were included in the review. For the 5 cases where an agreement could not be reached, a third reviewer was consulted for further assessment. Following this step, a total of 78 articles were included, and 77 articles were successfully retrieved for full-text screening. Sixteen studies were deemed ineligible based on the inclusion criteria and subsequently

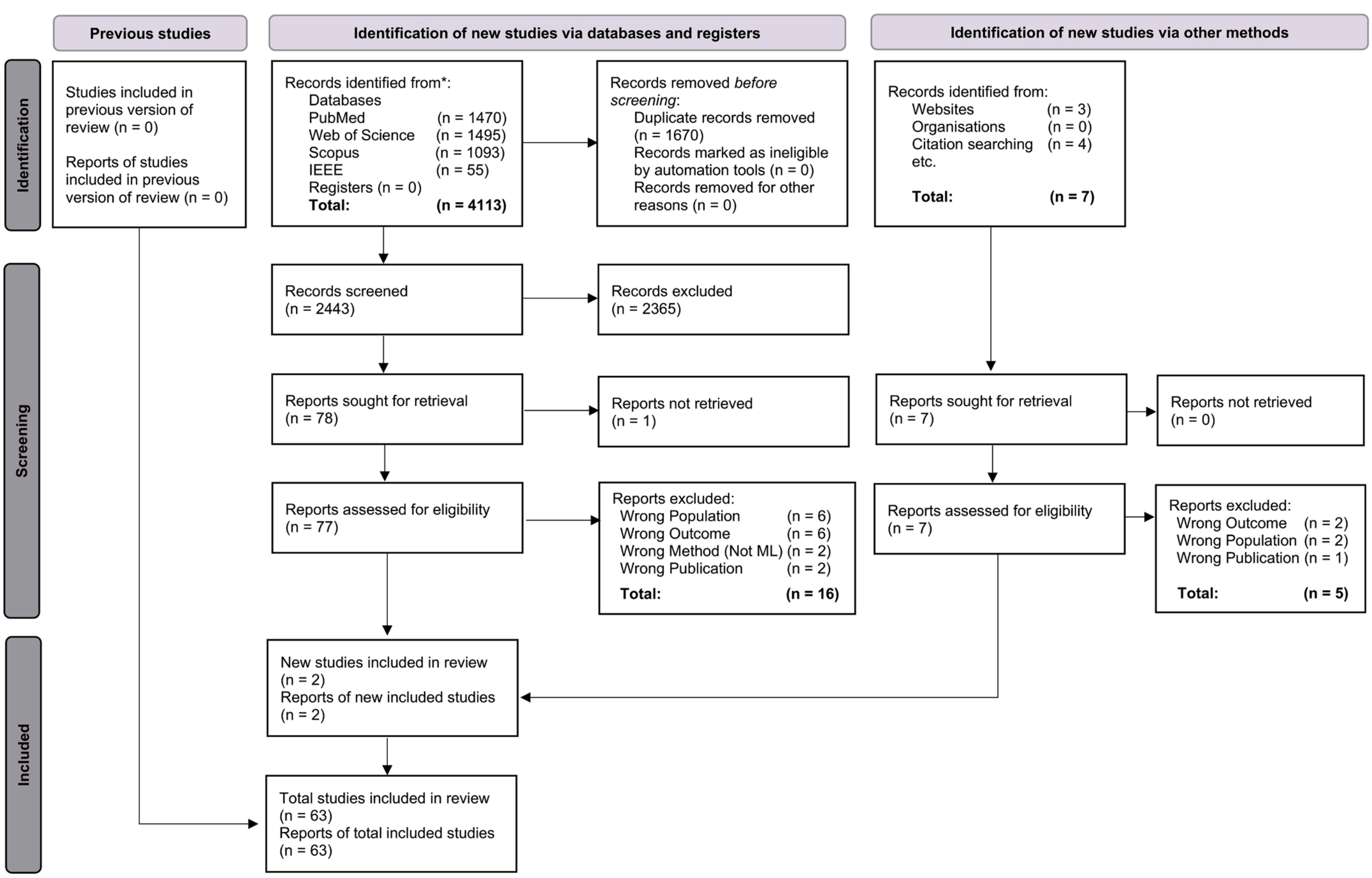

Fig. 1. PRISMA flow chart for paper selection.

excluded from the review. By adding 2 articles identified through hand searching, 63 unique articles were included in this review. Fig. 1 illustrates the PRISMA flow chart for the study selection process.

Out of the 63 included articles, 31 were evaluated using the questionnaire designed for supervised-based studies, while the remaining articles were assessed using the questionnaire for unsupervised-based studies. Three studies employed both supervised and unsupervised methods [37,45,46] and were categorized as supervised based on the main contribution and the employed validation methods.

3.1. Data extraction

Out of the 63 included articles, 37 focused on CP and 25 focused on stroke. Notably, in a study by Luo et al. [47], the pathology of included participants was not stated. Regarding the data acquisition methods, optoelectronic systems [25,26,28,31,33,37–40,45–66] and wearable sensors [30,67–75] are the primary standalone data sources and integrating the optoelectronic systems with force plates is the prevailing hybrid method for collecting gait data [34,66,76–83]. The most frequently used supervised methods in the included studies were artificial neural networks (ANN) (18 studies), support vector machine (SVM) (12 studies), and random forest (RF) (eight studies). Additionally, among unsupervised methods, k-means (19 studies), fuzzy clustering (six studies), and hierarchical clustering techniques (six studies) were commonly utilized (for more details, refer to the supplementary materials in Fig. S2). These ML algorithms are employed to analyze diverse range of standalone parameters, including kinematics, spatio-temporal, kinetics, and muscle activation and integrated parameters in which integration of spatio-temporal and kinematics is the most frequent (gait parameters frequency is illustrated in Fig. 3 in the supplementary materials). Table 3 summarizes relevant information extracted from the included studies. One of the most frequent objectives of applying these techniques was to investigate the different gait patterns within the same pathology [32–36,45,46,48,49,52–54,56,57,59–67,74–87]. This involves various applications, including severity classification of the disorders, classification of gait patterns based on established clinical benchmarks, and gait patterns clustering to uncover the significant parameters. Distinguishing between healthy and pathological gait was the next frequent objective in these studies [24–27,47,50,51,58,69–73, 88–90], which might contribute to the development of diagnostic tools and aids in the early detection and intervention of gait abnormalities in these individuals. Differentiating between various pathologies [28–31, 55,68], evaluating the efficacy of treatment modalities [37–39], and prescribing optimal treatment interventions [40] were additional applications of ML techniques in the analysis of gait patterns among individuals with CP and stroke. These applications aimed to improve clinical decision-making by providing objective measures of treatment outcomes, assisting in treatment planning, and monitoring progress, and facilitating personalized interventions based on each patient's specific needs and characteristics.

3.2. Quality assessment

3.2.1. Supervised ML methods

Among the 31 studies implementing supervised ML algorithms, 7 [30,51,57,66,68,82,83] were rated high-quality, 19 [24,25,27–29,37, 40,46,50,67,69–71,73,75,80,84,85,91] medium-quality, and 5 [38,39, 47,72,86] low-quality (Fig. 2a). All studies but 5 [38,39,50,70,85] demonstrated an appropriate study design and effectively communicated their research objectives (scoring details can be found in Table S1 in the supplementary materials).

3.2.1.1. Dataset. Information regarding the participants' inclusion criteria, demographic details, and clinical profiles was sufficiently reported in all studies but 9 [25,27,40,47,50,51,69,84,85]. Regarding the dataset size, 14 studies [24,29,30,37,39,46,51,67,68,70,71,75,80,86] included fewer than 50 subjects, and in 10 studies only [25,28,50,57,66, 69,82,84,85] participant count exceeded 100. The remaining studies included a number of subjects between 50 and 100 (for more details, refer to the supplementary materials in Fig. S4).

Sixteen papers adequately reported the data acquisition process, 7 studies [27,28,38–40,47,84] only a limited description, and 8 [39, 45–47,71,72,75,86] unclear definitions of training and test sets.

3.2.1.2. Feature engineering. Dimensionality reduction methods were used in 12 studies [24,30,40,51,57,67–70,73,80,84] to remove highly correlated features. All the studies except 7 [24,38,51,71,72,84,86] clearly explained the gait features selected as model input. Regarding providing explicit justification for the chosen features, only 1 research by Kim et al. [46] reported a convincing rationale.

3.2.1.3. Algorithm selection. Out of 31 studies, only 6 [7,12–14,19,23] provided adequate rationale for their algorithm choices. Computational cost, generalization capability, interpretability, and automatic feature extraction constituted the most frequent justifications in these studies. Other 7 studies mentioned the advantages and limitations of the selected ML method [40,51,66,71,82,83,91]. The outcomes of multiple ML algorithms were assessed across 18 studies [25,27,28,30,45,47,51,57, 66–69,71,80,82–85] to select the algorithm with the highest performance.

3.2.1.4. Algorithm design and validation. Most studies, except 7 [25,28, 38,69,72,80,86], detailed the design scheme. Notably, these explanations encompassed details such as the number of layers, neurons, activation functions, and optimization methods in ANN [25,38,45,67, 71–73,75,82,83,85]. Similarly, information about support vectors, kernels, and regularization parameters in SVM [30,37,47,50,51,69], as well as the number of trees in the RF methods [40,47,68] were among the

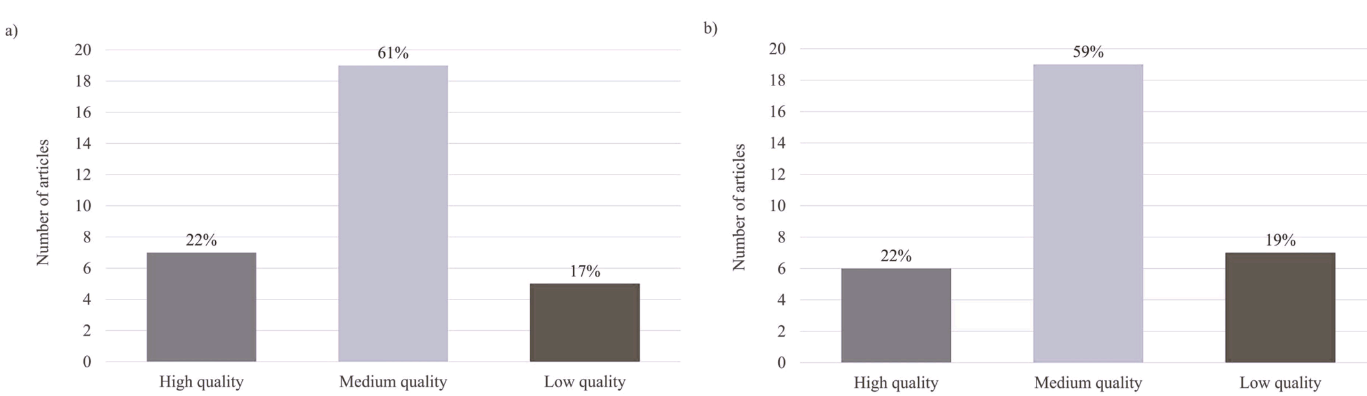

Fig. 2. Quality Assessment Scores for included articles. (a) Scores for studies with supervised methods. (b) Scores for studies with unsupervised methods.

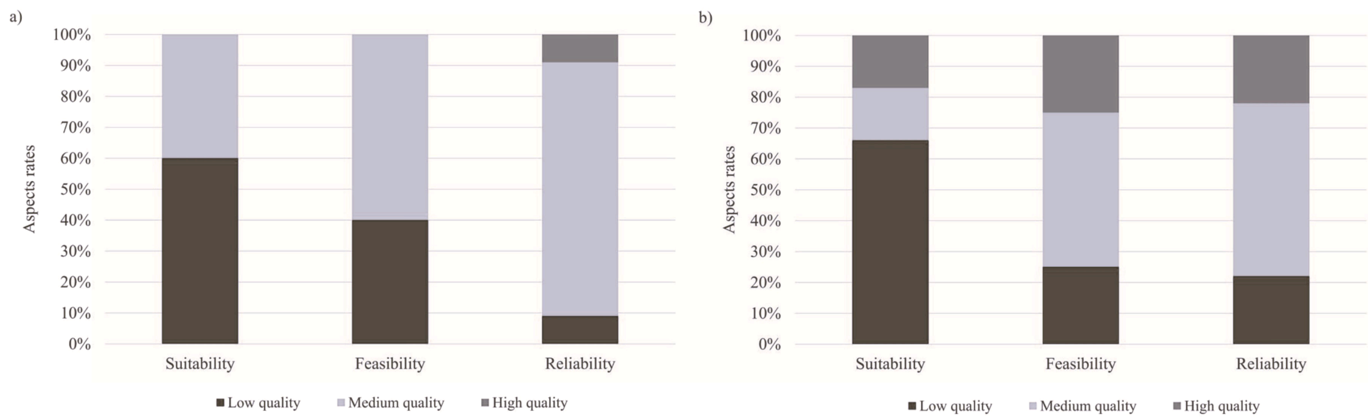

Fig. 3. Quality of addressing the methods' suitability, feasibility, and reliability. (a) Supervised algorithms. (b) Unsupervised algorithms.

most frequently presented elements in the studies. All but 6 studies [24,39,45,66,73,86] used suitable performance metrics for evaluating the results. In relation to the use of the appropriate techniques to deal with overfitting problem and provide more reliable estimates, 19 out of 31 employed cross-validation methods to evaluate their estimations [24,25,27–30,37,47,51,57,67–70,75,80,82,83,86] and the most of the studies except 10 [24,28,38,39,45–47,70,71,86] applied the independent test set to validate the generalization capacity of their proposed methods. In a broader context, it emerged that the findings derived from 12 studies [24,38,45,47,51,66,69,70,72,73,75,86] lacked sufficient support from the results.

In addition to analysing the key aspects of implementing ML algorithms, it was observed that the research limitations were reported in 17 studies [24,28,30,37,38,40,46,57,66,68–71,75,80,83,86]. Dataset size, diversity, and data quality were the most frequently addressed limitations. Lastly, regarding the contribution of the studies in adding value to the state of the art, 8 studies [27,30,67,68,71,82,83,85] received a full score.

No question related to suitability or feasibility received a high rating, whereas 9% of the questions concerning reliability were rated as high quality. In general, suitability and reliability received the lowest and highest scores, respectively. It is worth noting that the lack of justification for feature selection and for choice of ML algorithm, with only 3% and 19% of the studies achieving a full score, respectively, contributed to lowering the suitability and feasibility of the proposed methods. A schematic comprehensive depiction of how the questions related to the suitability, feasibility, and reliability of the proposed supervised methods were addressed is reported in Fig. 3a.

3.2.2. *Unsupervised ML methods*

Out of the 32 studies utilizing unsupervised ML algorithms, 6 [32,34,48,49,76,88] received a high score, 19 [26,31,33,35,52–56,58,59,61,62,65,74,77,79,81,90] received a medium score, and 7 [36,60,63,64,78,87,89] were rated low quality (Fig. 2b) (for more details about scoring refer to Table S2 in the supplementary materials).

All studies clearly defined the research objectives. However, the study design was not appropriately organized in 5 studies [49,59–61,89].

3.2.2.1. Dataset. All studies but 14 [33,35,36,56,58,60,61,63,64,74,78,79,87,89] presented detailed information regarding the characteristics and clinical profiles of the included participants. The number of subjects in the datasets ranged from 18 [78] to 2159 [36], and 23 studies [26,31,33–35,48,49,53–56,59,62–64,74,76–79,81,88,90] included fewer than 100 subjects (for more details, refer to the supplementary materials in Fig. S4); Hu et al. [89] did not report the sample size, and 4 studies [36,61,87,89] did not adequately explain the data acquisition framework.

3.2.2.2. Feature engineering. All papers but 3 [35,87,89] clearly explained the chosen gait features and pre-processing methods. Only 3 studies [48,58,65] reported a convincing rationale for selecting gait parameters. Furthermore, 7 studies clearly reported the use of automatic feature extraction methods: principal component analysis (PCA) [26,33,36,49], analysis of variance (ANOVA) [54], factorial analysis [56], and Pearson correlation [62].

3.2.2.3. Algorithm selection. Ten out of 32 studies provided clear justifications for the chosen ML algorithms [32,33,49,58,61,65,74,77,88,89]: assigning multiple memberships to the group in fuzzy clustering methods [32,33,58,65], finding the optimal number of clusters [49,61,88], evaluating the performance of new algorithms [74,77,89]. However, only 3 studies [48,53,54] discussed the advantages and drawbacks of the proposed algorithms, and none examined various methods to select the optimal one.

3.2.2.4. Algorithm design and validation. All studies but 8 [31,33,56,59,60,63,65,87] provided detailed information about algorithm design. All studies but 2 [78,89] clearly explained the methodology for determining the optimal number of clusters, but adequate clinical interpretation of the clusters was found in only 14 papers [26,32,34,35,48,49,52,53,55,56,76,79,87,88]. The validity of the obtained clusters was adequately assessed in all studies but 4 [48,60,64,89], where authors did not report any statistical method for evaluating the clusters.

In all studies but 11 [33,48,58,60,63–65,77,88–90], the results were reliable enough to support the conclusions. Moreover, only 4 studies demonstrated the efficacy of their findings to enhance clinical assessment [32,62,76,88].

Nineteen studies [26,31,34,35,48,49,52–56,58,60,61,63,76,79,88,90] addressed their limitations, small sample size and features limited to the sagittal plane motion being the most frequent. Nine studies [31,32,34,48,49,55,64,76,88] made significant contributions to the state of the art (see Table S2 in the supplementary material for more details).

For unsupervised methods, 17% of the questions related to suitability resulted in high-quality scores, while the rates were 25% for feasibility and 22% for reliability (Fig. 3b). Like studies on supervised methods, studies on unsupervised exhibited the lowest and highest scores for suitability and reliability, respectively. Notably, in both supervised and unsupervised methods, the rationale behind feature selection emerged as the most critical factor contributing to the reduction in suitability and feasibility, with only four [46,48,58,65] of the studies achieving a full score. On the other hand, the results indicated that studies employing unsupervised methods had a higher average rating when compared to those utilizing supervised algorithms.

Overall, from the quality assessment analysis, about 80% of the included studies were categorized as being of good quality (see Fig. 2). However, this positive rating primarily stemmed from high or very high

Table 3
Data extraction table for the 63 studies included in the review.

Authors	Pathologies	Main Objective	Sample Size	Measurement Tool	Gait Parameters	Study Plane	ML Algorithms	Group Numbers	Assessment Methods
Abbasi et al. [48]	Cerebral Palsy	Defining existing clusters of crouch gait patterns in children with spastic diplegic cerebral palsy	64	Eight infrared cameras of the Qualisys motion analysis system	Kinematics	Sagittal, Frontal, Transversal	Spars K-means	5	Observational evaluation using predefined criteria
Aguilera et al. [84]	Cerebral Palsy	Classification and discovery of gait patterns in children with spastic hemiplegia	278	Vicon, EMG	Kinematics, Kinetics, Muscle Activation	Sagittal, Frontal	Decision Tree, MLP Neural Networks, Via Regression, Adaptive Boosting	4	Accuracy, Specificity, Sensitivity, ROC
Amene et al. [49]	Cerebral Palsy	Identifying subgroups of children with pes planovalgus (PPV) secondary to CP	47	14-camera Vicon	Kinematics	Sagittal, Frontal, Transversal	K-means	6	ANOVA, post-hoc test
Armand et al. [32]	Cerebral Palsy	Clustering of sagittal ankle kinematic patterns	1376	Five-camera motion analysis system (Vicon1 VX), EMG, Force plates	Kinematics	Sagittal	Fuzzy c-means	3	ANOVA, post-hoc test
Bravo et al. [50]	Cerebral Palsy	Classification of the Spastic hemiplegia (SH) based on contralateral unaffected limb kinematics to differentiate from normal	933	Vicon 370 Motion System	Kinematics	Sagittal	Support Vector Machine	2	Specificity, Sensitivity
Burduk et al. [37]	Stroke	Clustering and classification of the stroke patients into 2 clusters to evaluate the efficiency of the rehabilitation method	50	Smartphone camera	Spatio-temporal	Not reported	K-means, Support Vector Machine, Quadratic Discriminant Analysis	2	Specificity, Sensitivity, ROC, Spearman's correlation coefficient, T-test, Shapiro Wilk
Carriero et al. [33]	Cerebral Palsy	Classifying spastic diplegic CP children into clinically recognizable groups according to their gait characteristics.	40	Seven infra-red cameras (Vicon 370)	Spatio-temporal, Kinematics	Sagittal, Frontal, Transversal	Fuzzy c-means	9	Assessed by applying new data
Chakraborty et al. [67]	Cerebral Palsy	Classification of pathological gait patterns using discrete wavelet and deep learning	18	IMU sensors	Kinematics	Not reported	Deep Neural Networks	2	Accuracy, Loss, Sensitivity, Positive Predictive Value, Negative Predictive Value, ROC
Chakraborty et al. [51]	Cerebral Palsy	Classification of CP and healthy subjects	40	Multiple Kinect Sensors (Slik F153)	Spatio-temporal	Sagittal	Extreme Learning Machine, MLP Neural Networks, K-Nearest Neighbor, Support Vector Machine	2	Accuracy, Specificity, Sensitivity
Chantraine et al. [52]	Stroke	Stiff knee severity classification in hemiparetic adults after stroke	115	Ten optoelectronic cameras (OQUS4, Qualisys AB)	Kinematics	Sagittal	K-means	5	ANOVA, post-hoc test, Kruskal-Wallis test

(continued on next page)

Table 3 (*continued*)

Authors	Pathologies	Main Objective	Sample Size	Measurement Tool	Gait Parameters	Study Plane	ML Algorithms	Group Numbers	Assessment Methods
Choisne et al. [53]	Cerebral Palsy	Classification of participants with cp and identifying the relation between orthotics type and gait pattern	98	Qualisys motion capture system with nine cameras (Qualisys AB)	Kinematics	Sagittal, Frontal, Transversal	Self-Organizing Map, K-means	6	ANOVA, Levene test of homogeneity of variance, Normality test
Cui et al. [24]	Stroke	Distinguishing the hemiparetic gait from normal gait and estimating the patient's lower limb motor function by a novel probability-based gait score	42	Six-Camera Qualisys motion capture system (Qualisys AB), EMG (Biomonitor ME6000, Mega Electronics), Bertec Force Plates	Kinematics, Ground Reaction Forces, Muscle Activation	Sagittal, Frontal, Transversal	Support Vector Machine, Neural Network, Random Forest, Naive Bayes, K-Nearest Neighbor, Fusion Algorithms (Classification-Based and Rule-Based)	2	Accuracy, Confusion Matrix
Darbandi et al. [54]	Cerebral Palsy	Classification of gait patterns in cp patients based on Rodda	84	Vicon Plug-in Gait	Kinematics	Sagittal	Fuzzy clustering	4	Accuracy, Specificity, Sensitivity, Positive Predictive Value, Negative Predictive Value
Dolatabadi et al. [88]	Stroke	Mixture-model clustering to spatiotemporal gait parameters to characterize the pathological gait pattern	88	GAITRite Force Plates	Spatio-temporal	Sagittal	Gaussian Mixture Model Clustering	3	Not Reported
Domagalska et al. [55]	Cerebral Palsy	Discrimination between children with unilateral cerebral palsy and mild scoliosis	96	3-D Real-time motion analysis system (CMS-HS 3D)	Spatio-temporal, Kinematics	Sagittal	K-means	3	ANOVA, post-hoc test
Domagalska–Szopa et al. [56]	Cerebral Palsy	Clustering of the gait patterns based on posture forms	58	3-D Real-time motion analysis system (CMS-HS 3D)	Spatio-temporal, Kinematics	Sagittal, Frontal, Transversal	K-means	4	ANOVA, post-hoc test
Ferrari et al. [85]	Cerebral Palsy	Classification of diplegic gait patterns	174	Optoelectronic cameras, EMG (No more info)	Kinematics	Sagittal, Frontal, Transversal	MLP Neural Networks, Recurrent Neural Networks (LSTM)	4	Accuracy, Confusion Matrix
Gestel et al. [66]	Cerebral Palsy	Classification of children with CP based on knee and ankle parameters	139	8 Infrared Vicon cameras, Force plates	Kinetics	Sagittal	Bayesian Network	4	Accuracy
Hsu et al. [68]	Stroke and other neurological disorders	Classification of participants with neurological disorders based on different configurations of wearable sensors	20	IMU sensors	Spatio-temporal, Kinematics	Not reported	Random Forest, Classifier, Decision Tree, Gaussian naïve Bayes, MLP Neural Network, AdaBoost	2	Accuracy, Sensitivity, Precision
Hu et al. [89]	Cerebral Palsy	Clustering the children with and without CP	Not reported	Not Reported	Not reported	Not reported	Mixture clustering Model	Not reported	Nor reported
Hussain et al. [69]	Stroke	Classification of Healthy and Stroke Participants based on muscular activity using myoelectric	123	EMG utilizing a Myoresearch DTS System	Muscle Activation	Not reported	Logistic Regression, Support Vector Machine, Decision Tree, MLP Neural Network,	2	Accuracy, Specificity, Sensitivity, Precision, Negative Predictive Value, ROC, Gini index

(*continued on next page*)

Table 3 (*continued*)

Authors	Pathologies	Main Objective	Sample Size	Measurement Tool	Gait Parameters	Study Plane	ML Algorithms	Group Numbers	Assessment Methods
		biomarkers evaluation					Discriminant Analysis Model		
Kaczmarczyk et al. [45]	Stroke	Classifying the gait patterns of post-stroke patients	74	3-DGA (Ariel Performance Analysis System (APAS))	Kinematics	Sagittal, Frontal	MLP Neural Networks, K-means, Discriminant Function Analysis	3	Accuracy, Coefficient of Determination, Confusion Matrix
Kamruzzaman et al. [25]	Cerebral Palsy	Discrimination between normal and CP children	156	Six-camera Vicon System	Spatio-temporal	Not reported	Support Vector Machine, MLP Neural networks, Linear Discriminant Analysis	2	Accuracy, Specificity, Sensitivity, ROC
Kim et al. [46]	Stroke	Identifying gait types of post-stroke hemiplegic patients directly from joint-level kinematics	36	Orthotrak motion capture system with eight cameras	Kinematics	Sagittal	Mini-batch K-means, K-means, Gaussian Mixture Model, Logistic Regression, Support Vector Machine, Random Forest, Gaussian Naive Bayes classifier	6	Accuracy, Sensitivity, Precision, F1-Score, ANOVA, Silhouette coefficient
Kim et al. [90]	Cerebral Palsy	Classification of muscle synergies in cp and healthy participants	28	Motion Capture Cameras (Vicon, Denver), EMG	Muscle Activation	Not reported	K-means	10	ANOVA, Spearman's Correlation coefficient, T-test
Kinsella et al. [34]	Stroke	Discrimination between gait patterns of stroke patients with equines deformity	23	Vicon Plug-in Gait Model, Force Plates (AMTI OR6–5)	Kinematics	Sagittal, Frontal	Hierarchical cluster analysis	3	ANOVA, post-hoc test
Krechowicz et al. [86]	Cerebral Palsy	Prediction of gait deviation indexes using only data extracted from the BS4P exam	29	BS4P, Isokinetic Dynamometer	Kinematics, Kinetics	Not reported	K-nearest Neighbor, Decision Tree Regression, Random Forest Regression, Gradient Boost Regression, Adaptive Boosting Regression	-	Coefficient of Determination
Krzak et al. [26]	Cerebral Palsy	Identifying clinically relevant subgroups among a sample of TD children and children with equinovarus due to hemiplegic CP	44	3-D motion analysis system and Vicon Nexus software	Kinematics	Sagittal, Frontal, Transversal	K-means	5	ANOVA, post-hoc test
Kuntze et al. [76]	Cerebral Palsy	Determining clusters of participants with CP based on multi-joint gait kinematics	37	8-camera optical motion analysis system (Motion Analysis), OR6–6 force plates (AMTI)	Kinematics	Sagittal, Frontal, Transversal	K-means	4	Silhouette coefficient
Laet et al. [57]	Cerebral Palsy	Classification of joint motion patterns for children with CP and studying the effect of expert knowledge in the supervised	356	Vicon Motion Systems	Kinematics	Sagittal, Frontal, Transversal	Naïve Bayes Classifier, Logestic Regression	3	Accuracy, F1-score,

(*continued on next page*)

Table 3 (*continued*)

Authors	Pathologies	Main Objective	Sample Size	Measurement Tool	Gait Parameters	Study Plane	ML Algorithms	Group Numbers	Assessment Methods
		classification system							
Lee et al. [70]	Stroke	Distinguish hemiplegic and normal gait	40	IMU sensors	Kinematics	Not reported	Random Forest	2	Accuracy, Specificity, Sensitivity, Positive Predictive Value
Iosa et al. [71]	Stroke	Classifying Stroke and healthy participants and identifying if the patient can return to the work	33	IMU sensors	Spatio-temporal, Kinematics	Sagittal, Frontal, Transversal	MLP Neural networks	2	Accuracy, Specificity, Sensitivity, ROC
Luo et al. [47]	Hemiplegia (The specific pathology was not reported)	Classification of participants with and without hemiplegia	60	Microsoft Kinect sensor	Spatio-temporal	Sagittal	Random Forest	2	Accuracy, ROC
MacWilliams et al. [28]	Cerebral Palsy	Discrimination between hereditary spastic paraplegia and cerebral palsy	706	Optoelectronic (No more info)	Spatio-temporal	Sagittal, Frontal, Transversal	Bayesian Additive Regression Trees	2	Specificity, Sensitivity, Confusion Matrix
Malley et al. [58]	Cerebral Palsy	Classification of the ambulation of neurologically intact children and those with cerebral palsy.	156	Six-Camera Vicon System	Spatio-temporal	Sagittal	Fuzzy clustering	5	Not reported
Manca et al. [59]	Stroke	Identification of foot-ankle complex dysfunction in gait patterns in hemiplegic patients	49	Six-camera motion analysis Vicon 460 system	Kinematics	Sagittal, Frontal, Transversal	Non-hierarchical cluster analysis	5	Post-hoc test, Kruskal-Wallis test
Mannini et al. [29]	Stroke, Huntington	Classification of different pathological gaits using probabilistic models	42	IMU sensors, Force Plates	Spatio-temporal	Not reported	Support Vector Machine	3	Mean Square Error, Confusion Matrix
Mathur et al. [72]	Stroke	Identification of different mobility levels of the entire group of patients with varying levels of disease	80	IMU sensor (Xsens Motion Capture System)	Spatio-temporal	Not reported	Logistic Regression, MLP Neural networks, Support vector machine, Extreme gradient boosting (XGBoost)	2	Sensitivity, Precision, F1-Score
Muhammad et al. [38]	Cerebral Palsy	Categorization of gait patterns to normal, pre-treatment, and post-treatment cp	55	3-DGA (Vicon Polygon software)	Spatio-temporal, Kinematics, Kinetics	Sagittal, Frontal, Transversal	MLP Neural networks	3	Accuracy, ROC
Mulroy et al. [35]	Stroke	Classification of gait patterns for stroke patients	52	Vicon motion analysis system, EMG, LIDO Active dynamometer Force Plates	Spatio-temporal, Kinematics	Sagittal	Non-hierarchical cluster analysis	4	Accuracy, ANOVA, Kruskal-Wallis test, Jack-Knife
O'Byrne et al. [60]	Cerebral Palsy	Categorization of gait patterns in CP participants	146	CODA-3 system	Kinematics	Sagittal	K-means	8	Not reported
Pauk et al. [77]	Stroke	Proposing a new biclustering method (KMB)	41	Motion tracking system (Motion Analysis Corp.),	Spatio-temporal	Not reported	agglomerative hierarchical clustering algorithm,	3	Not reported

(continued on next page)

Table 3 (*continued*)

Authors	Pathologies	Main Objective	Sample Size	Measurement Tool	Gait Parameters	Study Plane	ML Algorithms	Group Numbers	Assessment Methods
		in identifying three gait types		AMTI Force Plates			Biclustering algorithm (KMB)		
Pauk et al. [78]	Stroke	Clustering of gait patterns in the hemiplegia population	18	Motion tracking system (Motion Analysis Corp.), AMTI Force Plates	Kinetics	Not reported	Biclustering algorithm	Not reported	ANOVA, Mean Square Residue Score, Spearman's correlation coefficient
Prakash et al. [61]	Cerebral Palsy	Finding the optimal number of gait profiles using nature-based clustering algorithms	156	Six cameras of the Vicon system	Spatio-temporal	Not reported	K-means, GA, Hybrid GA, PSO, Hybrid-PSO, Fuzzy c-means	5, 4 (Differs for different algorithms)	Mean Square Error, Silhouette coefficient, T-test, Cluster purity index, Dunn Index.
Prokopowicz et al. [39]	Stroke	Comparing two post-stroke rehabilitation methods by machine learning approaches	50	Optoelectronic (No more info)	Spatio-temporal	Not reported	MLP Neural Networks	2	Not reported
Ries et al. [40]	Cerebral Palsy	Proposing a statistical orthosis selection model using the Random Forest Algorithm	1491	Optoelectronic (No more info)	Kinematics	Not reported	Random Forest	5	Accuracy, Specificity, Sensitivity, Positive Predictive Value, Negative Predictive Value, RMSE, Coefficient of determination, Matthews correlation coefficient
Roche et al. [62]	Cerebral Palsy	Determining principal gait patterns in adults with cp using the clustering approach	44	Motion Analysis System (Motion Analysis Corporation)	Spatio-temporal, Kinematics	Sagittal, Frontal	Hierarchical clustering method	5	ANOVA
Rozumalski et al. [36]	Cerebral Palsy	Clustering of CP who walk with excessive knee flexion at initial contact	2159	Not reported	Kinematics	Sagittal	K-means	5	Dunn Index
Sangeux et al. [87]	Cerebral Palsy	Classifying the sagittal gait patterns of patients with CP according to Rodda	776	Not reported	Kinematics	Sagittal	K-means	5	Accuracy, ANOVA, post-hoc test
Scheffer et al. [73]	Stroke	Distinguish between participants with hemiplegic stroke and healthy individuals	58	IMU sensors	Kinematics	Sagittal, Frontal, Transversal	MLP Neural Networks	2	Accuracy,
Sekiguchi et al. [79]	Stroke	Categorization of gait patterns after stroke based on ankle stiffness and use of ankle orthosis	79	Motion Analysis System with eight cameras (Motion Analysis Corporation), Force Plates	Kinematics	Sagittal	Hierarchical cluster analysis	3	ANOVA, post-hoc test, Kruskal-Wallis test
Straudi et al. [63]	Stroke	Clustering of hemiplegic gait in stroke patients	34	Vicon 460 motion analysis	Spatio-temporal, Kinematics	Sagittal	K-means	3	ANOVA
Sung et al. [80]	Stroke	Stroke severity classification based on	31	3-D motion capture cameras (Vicon Nexus)	Spatio-temporal, Kinematics, kinetics	Not reported	Support Vector Machine, Gradient Boosting,	3	Accuracy, Specificity, Sensitivity, Precision, F1-

(*continued on next page*)

Table 3 (*continued*)

Authors	Pathologies	Main Objective	Sample Size	Measurement Tool	Gait Parameters	Study Plane	ML Algorithms	Group Numbers	Assessment Methods
		walking symmetry					Decision Tree, Random Forest		Score, Confusion matrix
Szopa et al. [81]	Cerebral Palsy	Detecting gait patterns for unilateral CP participants using clustering	96	3-D real-time motion analysis (CMS-HS 3D), Force Plates	Spatio-temporal, Kinematics	Sagittal, Frontal, Transversal	K-means	3	ANOVA, post-hoc test
Tan et al. [74]	Stroke	Proposing a new Kinetic Index to characterize the gait deficits in stroke survivors and classifying stroke survivors based on TUG scores	30	IMU sensors, EMG	Spatio-temporal, Kinematics, Muscle Activation	Sagittal	Hierarchical Cluster Analysis	3	ANOVA
Toro et al. [64]	Cerebral Palsy	Clustering of different gait types in children with CP	67	Optoelectronic (No more info)	Kinematics	Sagittal	Hierarchical cluster analysis	13	Objective validation procedures based on clinical judgment
Vaughan et al. [65]	Cerebral Palsy	Introducing a simple gait nomogram based on the dynamic similarity hypothesis and monitoring of the functional status of CP children using Fuzzy clustering	747	Six-Camera Vicon system	Kinematics	Sagittal	Fuzzy Clustering	5	4 test data were used to test the validity of the proposed method
Wang et al. [30]	Stroke, Peripheral Neuropathy, Parkinson's	Classifying different pathologies by two shank-mounted IMUs	49	IMU sensors (InvenSense MPU-6050)	Spatio-temporal, Kinematics	Sagittal	Support Vector Machine	2	Accuracy, Specificity, Sensitivity, Confusion Matrix
Wang et al. [75]	Stroke	Detection and classification of stroke gaits as an aid to diagnosis and for application of appropriate rehabilitation methods selection	15	IMU sensors (APDM OPAL system)	Kinematics	Sagittal	Deep Neural Network	4	Accuracy, F1-Score
Wolf et al. [31]	Cerebral Palsy, Hereditary spastic Paraplegia	Distinguish between CP and HSP participants	87	Vicon 370 motion capture system	Kinematics	Sagittal	Fuzzy C-means	4, 5, 6	Not reported
Zhang et al. [82]	Cerebral Palsy	Classification of sagittal gait patterns for CP children with spastic diplegia	200	Eight-camera Vicon system, Force Plates	Kinematics	Sagittal	MLP Neural Network, Discriminant Analysis, Naive Bayes, Decision Tree, K-Nearest Neighbors, Support Vector Machine, Random Forest	4	Accuracy, Specificity, Sensitivity, ROC
Zhang et al. [27]	Cerebral Palsy	Distinguishing between CP and healthy participants	156	Six-camera Vicon system	Spatio-temporal	Not reported	Bayesian classification	2	Accuracy, Specificity, Sensitivity
Zwick et al. [83]	Cerebral Palsy	Classification method for equines deformity in spastic cerebral palsy. (Differentiate dynamic calf	66	Vicon 370, Force Plates	Kinematics, Kinetics	Not reported	Generalized Dynamic Neural Network	2	Specificity, Sensitivity, Positive Predictive Value, Negative Predictive Value

(*continued on next page*)

Table 3 (*continued*)

Authors	Pathologies	Main Objective	Sample Size	Measurement Tool	Gait Parameters	Study Plane	ML Algorithms	Group Numbers	Assessment Methods
		muscle tightness from fixed muscle contracture)							

scores in methodological aspects. At the same time, clinical-related criteria received deficient scores (see Fig. S1 in supplementary materials).

4. Discussion

This systematic review was conducted to address the existing gap between ML algorithms proposed in the literature to analyze or cluster GA data from children with CP and stroke patients and their clinical application.

4.1. Dataset

Data quality is pivotal in implementing a ML algorithm. The theory emphasizes the need for a dataset that adequately characterizes the statistical distribution of selected features in the chosen population including all relevant clinical information and reliable instrumental data.

Creating a sufficiently large and diverse dataset containing data collected for clinical GA with minimal noise and outliers is challenging. This complexity of providing datasets with adequate sample sizes and reliable data affects the feasibility, robustness, and reliability of ML methods.

One notable concern could be the use of synthetic data that may not accurately capture the intricacies of human movement dynamics and pathophysiological conditions, potentially limiting the generalizability of findings to real-world clinical settings. All the studies included in this review except one [89] used clinical datasets with a broad spectrum of sizes. In the study by Hu et al. [89] in which a mixed clustering method was proposed to analyze gait data of children with CP, there was a notable absence of information regarding the dataset utilized. This absence poses a challenge in discussing the results presented in the study. In studies utilizing supervised methods, the dataset sizes varied from 15 [75] to 933 [50], while for studies employing unsupervised algorithms, the range was broader, spanning from 18 [78] to 2159 [36] (see Table 3). Several studies utilized limited datasets, with 27% of the papers involving cohorts of 40 participants or fewer. This variety proved that no standard guideline for the ideal sample size definition exists. Reasonably, sample sizes depended on data availability without an a priori sample size design. Using small datasets can significantly impact the significance and generalizability of the results, i.e., of ML algorithms to make accurate predictions or classifications on unseen data.

An anecdotal example is the study by Kim et al. [90], in which 29 subjects were clustered into ten groups. In this case, aside from the increased risk of overfitting and the influence of noise and random variables in shaping clusters, extracting meaningful conclusions about the underlying patterns becomes challenging. On the other side of the spectrum, the research by Rozumalski et al. [36] featured the most extensive participant cohort among studies, encompassing 2159 individuals. Unfortunately, the authors of this study did not report critical factors such as demographic characteristics, the severity of functional limitations, and information regarding previous medical interventions. Noteworthy, approximately 40% of the assessed studies in this review did not report the patients' clinical characteristics (for more information see Fig. S1 in supplementary materials). These limitations can also compromise the clinical applicability of the results.

The reliability of proposed methods is closely tied to data acquisition, particularly regarding instrumental data quality limitations. In clinical practice, data acquired with both standalone and integrated methods can be affected by some operator-dependent errors, such as the misplacement of one or more markers (or inertial measurement units (IMU)) relative to the anatomical landmarks provided by the biomechanical model, differences in marker or IMU placement in successive evaluations, or e.g., in pre- and post-treatment studies, and the non-perfect identification of foot-contact or foot-off time instants [92,93]. These factors significantly influence the reliability of results, highlighting the critical importance of meticulous data acquisition techniques in ensuring robust and dependable outcomes. None of the studies addressed these practical considerations or implemented some pre-processing procedure to control for or decrease the effects of these factors in data. Besides this lack of data quality verification, only sagittal gait data was collected in 38% of studies (see Table 3). Restricting analysis to the sagittal plane may not adequately assess gait deviations and compensatory mechanisms that are typical for both post-stroke patients and - even more - children with CP [35,63,94].

The availability of large, diverse, and clean datasets is key in ML applications in clinical settings. Data collection should be carefully designed to improve model performance and reduce the risk of overfitting and considering real-world constraints like limited participant access, time limitations, ethical considerations, economic cost, and technological constraints. In addition, well-founded pre-processing techniques must be designed and applied to ensure input data quality.

4.2. Feature engineering

The choice of proper gait parameters, in alignment with the research objectives, plays a pivotal role in shaping the clinical relevance and suitability of algorithms. Employing pre-processing and feature extraction techniques can enhance the reliability of ML algorithms by reducing the influence of noise, outliers, and missing data and mitigating overfitting. Additionally, reducing the dimensionality of features enhances algorithm feasibility by reducing computational requirements.

Only 6% of the studies [46,48,58,65] addressed a well-defined rationale for the chosen parameters, which gave this item the lowest score in our assessment (see Fig. S1 in supplementary materials). The lack of clear justification for selected gait variables can affect the clinical utility and suitability of the methods employed. Making explicit the rationale behind the choice of parameters supports critical reasoning and also encourages the use of these variables by other researchers and clinicians who understand their significance and can translate them into daily clinical practice. Providing explicit reasoning behind the selected gait variables and using the clinically meaningful features can, therefore, enhance the relevance of the algorithm in clinical settings.

Regarding pre-processing methods, although many studies had implemented scaling methods, signal denoising was observed in only about 10% of these studies, with the low-pass Butterworth filter being the most used technique [29,51,80]. None of the studies provided a clear description of their approach to handling missing data. Since addressing these issues is critical for mitigating the shortcomings in data collection methods in clinical settings, the studies' limitations in implementing these techniques can pose significant practical challenges.

Finding the optimal number of features can be a challenging task in developing feature extraction and dimensionality reduction techniques. In two studies [58,65], the Authors employed a widely accepted heuristic to determine the appropriate number of features, where the number of subjects is chosen to be approximately ten times the dimension of the feature space [95]. Only in 6% of the studies did the number of samples and features follow this rule. As an example of the most extreme case, Mannini et al. [29] chose 18 features for a dataset comprising only 42 samples. This choice raises a concern related to the curse of dimensionality, where a substantial number of features in a constrained dataset leads to heightened computational complexities and an increased risk of overfitting in ML models [95]. The disparity between the dataset size and the number of features, mainly stemming from the constrained sizes of the datasets in various studies, can exert notable effects on the interpretability, computational burden, and performance of ML methods.

4.3. Algorithm selection

Algorithm interpretability is a significant consideration in determining the suitability of the selected ML method for clinical applications. Simultaneously, selecting an algorithm with better performance indices can be more reliable. Moreover, the feasibility of the algorithms in the real-world setting can be affected by their computational cost.

Clinicians prefer to understand how the algorithm makes predictions and identifies specific gait patterns. Therefore, interpretable methods are preferable for clinical applications and can potentially cultivate trust among clinicians and promote their acceptance and utilization. Only one study by Gestel et al. [66] explicitly highlighted the interpretability of their chosen method as a key rationale in algorithm selection. In their study, a Bayesian network approach was used to classify children with CP. Inherent limitations of some ML algorithms in interpretability could be the main reason for this limitation in studies. The shortening of interpretability in the proposed methods constitutes a critical challenge to the clinical utility of these models. Computational requirement, another factor in selecting an ML algorithm, was noted only in one study [51]. The selection of ML algorithms is directly linked to the feasibility of methods, particularly concerning computational requirements. Given the cost constraints and hardware limitations often present in clinical settings, considering computational requirements can significantly enhance the feasibility of the methods in real-world settings. This consideration ensures that the chosen algorithms are not only effective but also practical for implementation in real-world clinical scenarios. Finally, to attain satisfactory performance, typically the primary objective in the development of ML techniques, the optimal selection of an algorithm plays a critical role. To achieve this goal, assessing the outcomes of several chosen methods to reach the best solution can be a proper approach. This procedure was the main criterion in selecting the methods in the assessed literature.

When choosing an appropriate ML method for clinical applications, it is essential to take all these factors into account. However, there exists a trade-off between these aspects. For instance, enhancing performance may lead to increased computational demands. Therefore, evaluating the pros and cons of ML algorithms and selecting the algorithm according to the research objective and practical limitations becomes crucial. Ten studies [40,45,48,51,53,54,66,71,82,83] addressed the advantages and drawbacks of developed ML methods, while no one addressed practical challenges regarding developed methods. Thoughtful evaluation of these aspects, aligned with research objectives and real-world constraints, plays a paramount role in mitigating the inherent limitations and optimizing the effectiveness of these methods in clinical settings.

4.4. Algorithm design and validation

Developing appropriate design and validation methods holds substantial importance in ensuring the clinical reliability and feasibility of the proposed methods. Furthermore, in unsupervised methods, accurately determining the number of clusters can significantly improve the clinical relevance of the results, thereby increasing the suitability of the algorithms for clinical applications.

All included studies utilized the random sample selection approaches for constructing test sets. In the research conducted by Lee et al. [70], the authors employed 80 sets of gait data collected from 40 individuals for training an RF algorithm. To create the training and test sets, they randomly divided the data, allocating 75% for training and 25% for testing. Therefore, because of the random allocation of various features from each subject, the data from each participant could be present in both the training and testing datasets. This non-independence of the test set raises concerns about the reliability of the reported 100% scores for performance metrics in this study. It became evident that the results presented in the paper were highly prone to overfitting. While random test set selection is essential to provide an unbiased evaluation of a model's performance, it cannot ensure the representativeness of the test set, which is imperative for a more efficient model. To generate a representative test set, stratified sampling is a noteworthy strategy, a statistical technique ensuring that each class is proportionally represented in the test set [96]. This method was not addressed in the studies, utilizing which can improve the reliability of assessments.

In most studies, applying an independent test set and cross-validation methods were two frequent considerations for handling overfitting and performing a realistic evaluation of the proposed methods (see Fig. S1 in supplementary materials). Constructing an appropriate test set and developing cross-validation methods are crucial in implementing a reliable algorithm. However, it is important to note that cross-validation techniques can affect the feasibility of the algorithms by increasing computational and timing requirements. Therefore, balancing these factors is essential to achieve the optimal solution based on clinical objectives.

Effective hyperparameter tuning enhances the method's reliability. Most of the studies represented the details of the proposed ML algorithms. In this regard, selecting the optimal number of clusters in unsupervised methods poses challenges. Despite most of the studies that conducted mathematical methods, in eight studies [54–56,60,63,81,87,90], the experts' knowledge was employed to predefine the number of clusters. Although employing mathematical methods to determine the number of clusters is effective in revealing hidden patterns within gait data, it may lack the clinical explainability that is a primary advantage of expert-defined clusters. In the study by Mulroy et al. [35], the optimal number of clusters was determined by evaluating the resultant clinical characteristics of a practical range of group numbers and by maximizing the R ratio. This approach combines mathematical techniques with expert insights, enabling not only the identification of concealed patterns through algorithms but also making the findings more clinically comprehensible. Therefore, the utilization of such integrated approaches can be more effective for clinical applications.

While most of the studies employed robust validation techniques to evaluate the performance of proposed algorithms, five studies with supervised methods [24,39,45,66,73] relied solely on accuracy as the metric for assessing their results. Considering the critical importance of false negatives and false positives in clinical decision-making, depending solely on accuracy may prove inadequate for ensuring the reliability of a ML algorithm. In most of the studies utilizing unsupervised methods, statistical approaches were applied to evaluate the quality of the clusters. Applying these techniques can assess the consistency of the clusters. On the other hand, two studies [48,64] exclusively relied on clinical judgments and predefined clinical criteria for assessing cluster quality. While utilizing expert knowledge to assess the findings can effectively enhance their clinical relevance, the consistency of the clusters could be adversely affected by a lack of statistical analysis methods. Sangeux et al. [87] conducted an integrated approach. They not only utilized statistical methods such as ANOVA and post-hoc tests to evaluate cluster quality but also validated their findings by comparing them to the results of classifications conducted by experts. Employing methods that integrate expert knowledge with statistical techniques can effectively ensure both the reliability and clinical relevance of the findings.

Our analysis highlighted that adapting ML algorithms for clinical applications is mainly challenging in terms of suitability, primarily due to the lack of justification in selecting both features and ML algorithms. Additionally, another notable issue that affected the suitability, especially in the realm of unsupervised methods, was the difficulty in achieving clinical interpretability of the results. In these algorithms, determining the optimal number of clusters emerged as a crucial factor for ensuring that the clinical outcomes can be readily understood and explained.

With an average of around 70% of studies achieving medium and high quality, it is worth noting that the feasibility of the methods was reasonably well-considered. Nevertheless, it is important to highlight that the feasibility was still affected by the inadequate justification provided for selecting features and algorithms.

The reliability of the systems gained the highest score, indicating that many researchers prioritized accuracy in their systems over their applicability and usability in real-world clinical contexts. Nevertheless, it is crucial to emphasize that focusing solely on the reliability of the systems is insufficient for achieving clinical applicability, as addressing real-world constraints characterized by the suitability and feasibility of these systems is of paramount importance.

A significant diversity among study protocols was evident in various aspects, including participant demographics, data acquisition methods, extracted features, ML algorithms, and validation techniques. This diversity reflects the multifaceted nature of clinical gait analysis and the complexity of applying ML algorithms to this domain.

For instance, different studies focused on different populations, utilized various types of sensors, and employed a range of ML algorithms. The impact of this diversity on the analysis can be profound. Variations in participant characteristics, such as age, severity of condition, and comorbidities, can introduce heterogeneity that may influence the generalizability of findings across different patient populations. Differences in data acquisition methods may affect the quality and reliability of the input data, potentially influencing the performance of ML algorithms. Moreover, the choice of ML algorithms and validation techniques can significantly impact the interpretability and generalizability of results.

The findings of studies included in this study may not have been translated directly to the clinical applications. However, they can contribute to enhance our understanding of gait patterns, identifying relevant features, and exploring predictive models using various algorithms.

5. Conclusion

Although ML algorithms are powerful tools for handling vast and complex gait data, our current review has highlighted a notable lack of clinical relevance in the ML methods proposed for analysing clinical gait data from individuals with CP and stroke. The overall methodological quality of the evaluated studies regarding their appropriateness for clinical applications was found to be low. Most of the studies classified the samples into some classes focusing on enhancing the specified performance indices. However, the reliability and practicality of these methods for clinical settings were compromised due to insufficient justification for selecting features and ML algorithms and using datasets that did not adequately represent the population. Furthermore, there was a notable absence of explanations regarding how the resulting clusters were related to the impairments observed in individuals with CP and stroke, particularly in unsupervised ML methods. To develop ML methods that are genuinely applicable in clinical settings, it is essential to consider various aspects such as dataset quality, feature engineering, model selection, and interpretability of the results more comprehensively. The findings in this review can be used to develop more robust gait data analysis methods using ML algorithms that are better equipped to address the constraints and complexities of real-world clinical scenarios.

CRediT authorship contribution statement

Farshad Samadi Kohnehshahri: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Andrea Merlo:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Davide Mazzoli:** Writing – review & editing, Supervision, Conceptualization. **Maria Chiara Bò:** Writing – original draft, Data curation. **Rita Stagni:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization.

Funding

This study was equally funded by the European Union Next Generation EU-PNRR program and by the Sol et Salus Hospital, Rimini, Italy. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was equally funded by 1) PNRR DM 352/2022 funded by European Union - NextGenerationEU and 2) the Sol et Salus Hospital, Rimini, Italy.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.gaitpost.2024.04.007.

References

[1] Y. Celik, S. Stuart, W.L. Woo, A. Godfrey, Gait analysis in neurological populations: Progression in the use of wearables, Med. Eng. Phys. 87 (2021) 9–29.

[2] S. Armand, G. Decoulon, A. Bonnefoy-Mazure, Gait analysis in children with cerebral palsy, EFORT Open Rev. 1 (12) (2016) 448–460.

[3] T. Truelsen, B. Piechowski-Jóźwiak, R. Bonita, C. Mathers, J. Bogousslavsky, G. Boysen, Stroke incidence and prevalence in Europe: a review of available data, Eur. J. Neurol. 13 (6) (2006) 581–598.

[4] T.A. Wren, K.J. Elihu, S. Mansour, S.A. Rethlefsen, D.D. Ryan, M.L. Smith, et al., Differences in implementation of gait analysis recommendations based on affiliation with a gait laboratory, Gait Posture 37 (2) (2013) 206–209.

[5] T.A. Wren, C.A. Tucker, S.A. Rethlefsen, G.E. Gorton III, S. Õunpuu, Clinical efficacy of instrumented gait analysis: systematic review 2020 update, Gait Posture 80 (2020) 274–279.

[6] I. Campanini, M. Cosma, M. Manca, A. Merlo, Added value of dynamic EMG in the assessment of the equinus and the equinovarus foot deviation in stroke patients and barriers limiting its usage, Front. Neurol. 11 (2020) 583399.

[7] T.A. Wren, G.E. Gorton III, S. Ounpuu, C.A. Tucker, Efficacy of clinical gait analysis: a systematic review, Gait Posture 34 (2) (2011) 149–153.

[8] M.W. Whittle, Clinical gait analysis: A review, Hum. Mov. Sci. 15 (3) (1996) 369–387.

[9] D. Mazzoli, E. Giannotti, M. Manca, M. Longhi, P. Prati, M. Cosma, et al., Electromyographic activity of the vastus intermedius muscle in patients with stiff-knee gait after stroke. A retrospective observational study, Gait Posture 60 (2018) 273–278.

[10] D.A. Fuller, M.A.E. Keenan, A. Esquenazi, J. Whyte, N.H. Mayer, R. Fidler-Sheppard, The impact of instrumented gait analysis on surgical planning: treatment of spastic equinovarus deformity of the foot and ankle, Foot ankle Int. 23 (8) (2002) 738–743.

[11] T.A. Wren, N.Y. Otsuka, R.E. Bowen, A.A. Scaduto, L.S. Chan, M. Sheng, et al., Influence of gait analysis on decision-making for lower extremity orthopaedic surgery: Baseline data from a randomized controlled trial, Gait Posture 34 (3) (2011) 364–369.

[12] R. Baker, A. Esquenazi, M.G. Benedetti, K. Desloovere, Gait analysis: clinical facts, Eur. J. Phys. Rehabil. Med 52 (4) (2016) 560–574.

[13] A. Merlo, I. Campanini, Impact of instrumental analysis of stiff knee gait on treatment appropriateness and associated costs in stroke patients, Gait Posture 72 (2019) 195–201.
[14] J. Park, M.K. Seeley, D. Francom, C.S. Reese, J.T. Hopkins, Functional vs. traditional analysis in biomechanical gait data: an alternative statistical approach, J. Hum. Kinet. 60 (1) (2017) 39–49.
[15] A. Duhamel, J. Bourriez, P. Devos, P. Krystkowiak, A. Destee, P. Derambure, et al., Statistical tools for clinical gait analysis, Gait Posture 20 (2) (2004) 204–212.
[16] M.H. Schwartz, A. Rozumalski, The gait deviation index: a new comprehensive index of gait pathology, Gait Posture 28 (3) (2008) 351–357.
[17] R. Baker, J.L. McGinley, M.H. Schwartz, S. Beynon, A. Rozumalski, H.K. Graham, et al., The gait profile score and movement analysis profile, Gait Posture 30 (3) (2009) 265–269.
[18] I. Campanini, A. Merlo, Reliabilty, smallest real difference and concurrent validity of indices computed from GRF components in gait of stroke patients, Gait Posture 30 (2) (2009) 127–131.
[19] A. Phinyomark, G. Petri, E. Ibáñez-Marcelo, S.T. Osis, R. Ferber, Analysis of big data in gait biomechanics: Current trends and future directions, J. Med. Biol. Eng. 38 (2018) 244–260.
[20] L. Lee, W.E.L. Grimson, Gait analysis for recognition and classification. in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, IEEE, 2002.
[21] E. Dolatabadi, B. Taati, A. Mihailidis, An automated classification of pathological gait using unobtrusive sensing technology, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (12) (2017) 2336–2346.
[22] R. Best, R. Begg, Overview of movement analysis and gait features. in *Computational intelligence for movement sciences: neural networks and other emerging techniques*, IGI Global, 2006, pp. 1–69.
[23] P. Khera, N. Kumar, Role of machine learning in gait analysis: a review, J. Med. Eng. Technol. 44 (8) (2020) 441–467.
[24] C. Cui, G.-B. Bian, Z.-G. Hou, J. Zhao, G. Su, H. Zhou, et al., Simultaneous recognition and assessment of post-stroke hemiparetic gait by fusing kinematic, kinetic, and electrophysiological data, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (4) (2018) 856–864.
[25] J. Kamruzzaman, R.K. Begg, Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait, IEEE Trans. Biomed. Eng. 53 (12) (2006) 2479–2490.
[26] J.J. Krzak, D.M. Corcos, D.L. Damiano, A. Graf, D. Hedeker, P.A. Smith, et al., Kinematic foot types in youth with equinovarus secondary to hemiplegia, Gait Posture 41 (2) (2015) 402–408.
[27] B.-l Zhang, Y. Zhang, R.K. Begg, Gait classification in children with cerebral palsy by Bayesian approach, Pattern Recognit. 42 (4) (2009) 581–586.
[28] B.A. MacWilliams, K.L. Carroll, A.K. Stotts, L.M. Kerr, M.H. Schwartz, Discrimination between hereditary spastic paraplegia and cerebral palsy based on gait analysis data: a machine learning approach, Gait Posture 98 (2022) 34–38.
[29] A. Mannini, D. Trojaniello, A. Cereatti, A.M. Sabatini, A machine learning framework for gait classification using inertial sensors: application to elderly, post-stroke and Huntington's disease patients, Sensors 16 (1) (2016).
[30] L. Wang, Y. Sun, Q. Li, T. Liu, J. Yi, Two shank-mounted IMUs-based gait analysis and classification for neurological disease patients, IEEE Robot. Autom. Lett. 5 (2) (2020) 1970–1976.
[31] S.I. Wolf, F. Braatz, D. Metaxiotis, P. Armbrust, T. Dreher, L. Döderlein, et al., Gait analysis may help to distinguish hereditary spastic paraplegia from cerebral palsy, Gait Posture 33 (4) (2011) 556–561.
[32] S. Armand, E. Watelain, M. Mercier, G. Lensel, F.-X. Lepoutre, Identification and classification of toe-walkers based on ankle kinematics, using a data-mining method, Gait Posture 23 (2) (2006) 240–248.
[33] A. Carriero, A. Zavatsky, J. Stebbins, T. Theologis, S.J. Shefelbine, Determination of gait patterns in children with spastic diplegic cerebral palsy using principal components, Gait Posture 29 (1) (2009) 71–75.
[34] S. Kinsella, M. Kieran, Gait pattern categorization of stroke participants with equinus deformity of the foot, Gait & posture 27 (1) (2008) 144–151.
[35] S. Mulroy, J. Gronley, W. Weiss, C. Newsam, J. Perry, Use of cluster analysis for gait pattern classification of patients in the early and late recovery phases following stroke, Gait Posture 18 (1) (2003) 114–125.
[36] A. Rozumalski, M.H. Schwartz, Crouch gait patterns defined using k-means cluster analysis are related to underlying clinical pathology, Gait Posture 30 (2) (2009) 155–160.
[37] R. Burduk, I. Rojek, E. Mikołajewska, D. Mikołajewski, Post-stroke gait classification based on feature space transformation and data labeling, Appl. Sci. 12 (22) (2022) 11346.
[38] J. Muhammad, S. Gibbs, R. Abboud, S. Anand, W. Wang, Gait pattern recognition in cerebral palsy patients using neural network modelling, J. Ayub Med. Coll., Abbottabad: JAMC 27 (4) (2015) 754–758.
[39] P. Prokopowicz, D. Mikołajewski, K. Tyburek, E. Mikołajewska, Computational gait analysis for post-stroke rehabilitation purposes using fuzzy numbers, fractal dimension and neural networks, Bull. Pol. Acad. Sci.: Tech. Sci. (2) (2020).
[40] A.J. Ries, T.F. Novacheck, M.H. Schwartz, A data driven model for optimal orthosis selection in children with cerebral palsy, Gait Posture 40 (4) (2014) 539–544.
[41] P. Rajpurkar, E. Chen, O. Banerjee, E.J. Topol, AI in health and medicine, Nat. Med. 28 (1) (2022) 31–38.
[42] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, Int. J. Surg. 88 (2021) 105906.
[43] J. Fayad, P.E. Eltes, A. Lazary, L. Cristofolini, R. Stagni, Stereophotogrammetric approaches to multi-segmental kinematics of the thoracolumbar spine: a systematic review, BMC Musculoskelet. Disord. 23 (1) (2022) 1080.
[44] R. Caldas, M. Mundt, W. Potthast, F.B. de Lima Neto, B. Markert, A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms, Gait Posture 57 (2017) 204–210.
[45] K. Kaczmarczyk, A. Wit, M. Krawczyk, J. Zaborski, Gait classification in post-stroke patients using artificial neural networks, Gait Posture 30 (2) (2009) 207–210.
[46] H. Kim, Y.-H. Kim, S.-J. Kim, M.-T. Choi, Pathological gait clustering in post-stroke patients using motion capture data, Gait Posture 94 (2022) 210–216.
[47] G. Luo, Y. Zhu, R. Wang, Y. Tong, W. Lu, H. Wang, Random forest–based classsification and analysis of hemiplegia gait using low-cost depth cameras. Med. Biol. Eng. Comput. 58 (2020) 373–382.
[48] L. Abbasi, Z. Rojhani-Shirazi, M. Razeghi, H. Raeisi-Shahraki, Kinematic cluster analysis of the crouch gait pattern in children with spastic diplegic cerebral palsy using sparse K-means method, Clin. Biomech. 81 (2021) 105248.
[49] J. Amene, J.J. Krzak, K.M. Kruger, L. Killen, A. Graf, H. Altiok, et al., Kinematic foot types in youth with pes planovalgus secondary to cerebral palsy, Gait Posture 68 (2019) 430–436.
[50] R.J. BRAVO, O.C.D.E. CASTRO, A.J. SALAZAR, Spastic hemiplegia gait characterization using support vector machines: contralateral lower limb, Rev. De. la Fac. De. Ing. ía Univ. Cent. De. Venez. 21 (2) (2006) 111–119.
[51] S. Chakraborty, A. Nandy, Automatic diagnosis of cerebral palsy gait using computational intelligence techniques: a low-cost multi-sensor approach, IEEE Trans. Neural Syst. Rehabil. Eng.: a Publ. IEEE Eng. Med. Biol. Soc. 28 (11) (2020) 2488–2496.
[52] F. Chantraine, C. Schreiber, J.A.C. Pereira, J. Kaps, F. Dierick, Classification of stiff-knee gait kinematic severity after stroke using retrospective k-means clustering algorithm, J. Clin. Med. 11 (21) (2022).
[53] J. Choisne, N. Fourrier, G. Handsfield, N. Signal, D. Taylor, N. Wilson, et al., An unsupervised data-driven model to classify gait patterns in children with cerebral palsy, J. Clin. Med. 9 (5) (2020) 1432.
[54] H. Darbandi, M. Baniasad, S. Baghdadi, A. Khandan, A. Vafaee, F. Farahmand, Automatic classification of gait patterns in children with cerebral palsy using fuzzy clustering method, Clin. Biomech. 73 (2020) 189–194.
[55] M. Domagalska-Szopa, A. Szopa, Gait pattern differences between children with mild scoliosis and children with unilateral cerebral palsy, PLoS One 9 (8) (2014) e103095.
[56] M. Domagalska-Szopa, A. Szopa, Gait pattern differences among children with bilateral cerebral palsy, Front. Neurol. 10 (2019) 183.
[57] T. De Laet, E. Papageorgiou, A. Nieuwenhuys, K. Desloovere, Does expert knowledge improve automatic probabilistic classification of gait joint motion patterns in children with cerebral palsy? PLoS One 12 (6) (2017) e0178378.
[58] M.J. O'Malley, M.F. Abel, D.L. Damiano, C.L. Vaughan, Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters, IEEE Trans. Rehabil. Eng. 5 (4) (1997) 300–309.
[59] M. Manca, G. Ferraresi, M. Cosma, L. Cavazzuti, M. Morelli, M. Benedetti, Gait patterns in hemiplegic patients with equinus foot deformity, BioMed. Res. Int. 2014 (2014).
[60] J.M. O'Byrne, A. Jenkinson, T. O'brien, Quantitative analysis and classification of gait patterns in cerebral palsy using a three-dimensional motion analyzer, J. Child Neurol. 13 (3) (1998) 101–108.
[61] C. Prakash, R. Kumar, N. Mittal, Optimized clustering techniques for gait profiling in children with cerebral palsy for rehabilitation, Comput. J. 61 (11) (2018) 1683–1694.
[62] N. Roche, D. Pradon, J. Cosson, J. Robertson, C. Marchiori, R. Zory, Categorization of gait patterns in adults with cerebral palsy: a clustering approach, Gait Posture 39 (1) (2014) 235–240.
[63] S. Straudi, M. Manca, E. Aiello, G. Ferraresi, S. Cavazza, N. Basaglia, Sagittal plane kinematic analysis of the six-minute walk test: a classification of hemiplegic gait, Eur. J. Phys. Rehabil. Med 45 (3) (2009) 341–347.
[64] B. Toro, C.J. Nester, P.C. Farren, Cluster analysis for the extraction of sagittal gait patterns in children with cerebral palsy, Gait Posture 25 (2) (2007) 157–165.
[65] C.L. Vaughan, M.J. O'Malley, A gait nomogram used with fuzzy clustering to monitor functional status of children and young adults with cerebral palsy, *Dev. Med. Child Neurol.* (2005) 377–383 (England).
[66] L. Van Gestel, T. De Laet, E. Di Lello, H. Bruyninckx, G. Molenaers, A. Van Campenhout, et al., Probabilistic gait classification in children with cerebral palsy: A Bayesian approach, Res. Dev. Disabil. 32 (6) (2011) 2542–2552.
[67] J. Chakraborty, A. Nandy, Discrete wavelet transform based data representation in deep neural network for gait abnormality detection, Biomed. Signal Process. Control 62 (2020) 102076.
[68] W.-C. Hsu, T. Sugiarto, Y.-J. Lin, F.-C. Yang, Z.-Y. Lin, C.-T. Sun, et al., Multiple-wearable-sensor-based gait classification and analysis in patients with neurological disorders, Sensors 18 (10) (2018) 3397.
[69] I. Hussain, S.-J. Park, Prediction of myoelectric biomarkers in post-stroke gait, Sensors 21 (16) (2021) 5334.
[70] J. Lee, S. Park, H. Shin, Detection of hemiplegic walking using a wearable inertia sensing device, Sensors 18 (6) (2018) 1736.
[71] M. Iosa, E. Capodaglio, S. Pelà, B. Persechino, G. Morone, G. Antonucci, et al., Artificial Neural Network Analyzing Wearable Device Gait Data for Identifying Patients With Stroke Unable to Return to Work, Front. Neurol. 12 (2021) 650542.
[72] D. Mathur, D. Bhatia, Gait classification of stroke survivors-An analytical study, J. Interdiscip. Math. 25 (1) (2022) 163–181.
[73] C. Scheffer, T. Cloete, 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. Eng. 15 (3) (2012) 285–294.
[74] M.-G. Tan, J.-H. Ho, H.-T. Goh, H.K. Ng, L.A. Latif, M. Mazlan, A new fractal-based kinetic index to characterize gait deficits with application in stroke survivor functional mobility assessment, Biomed. Signal Process. Control 52 (2019) 403–413.
[75] F.-C. Wang, S.-F. Chen, C.-H. Lin, C.-J. Shih, A.-C. Lin, W. Yuan, et al., Detection and classification of stroke gaits by deep neural networks employing inertial measurement units, Sensors 21 (5) (2021) 1864.
[76] G. Kuntze, A. Nettel-Aguirre, G. Ursulak, I. Robu, N. Bowal, S. Goldstein, et al., Multi-joint gait clustering for children and youth with diplegic cerebral palsy, PLoS One 13 (10) (2018) e0205174.
[77] J. Pauk, K. Minta-Bielecka, Gait patterns classification based on cluster and bicluster analysis, Biocybern. Biomed. Eng. 36 (2) (2016) 391–396.
[78] J. Pauk, K. Minta-Bielecka, A new classification of hemiplegia gait patterns based on bicluster analysis of joint moments, Acta Bioeng. Biomech. 18 (4) (2016).
[79] Y. Sekiguchi, K. Honda, D. Owaki, S.-I. Izumi, Classification of ankle joint stiffness during walking to determine the use of ankle foot orthosis after stroke, Brain Sci. 11 (11) (2021) 1512.
[80] J. Sung, S. Han, H. Park, S. Hwang, S.J. Lee, J.W. Park, et al., Classification of Stroke severity using clinically relevant symmetric gait features based on recursive feature elimination with cross-validation, IEEE Access 10 (2022) 119437–119447.
[81] A. Szopa, M. Domagalska-Szopa, A. Czamara, Gait pattern differences in children with unilateral cerebral palsy, Res. Dev. Disabil. 35 (10) (2014) 2261–2266.
[82] Y. Zhang, Y. Ma, Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia, Comput. Biol. Med. 106 (2019) 33–39.
[83] E.B. Zwick, L. Leistritz, B. Milleit, V. Saraph, G. Zwick, M. Galicki, et al., Classification of equinus in ambulatory children with cerebral palsy-discrimination between dynamic tightness and fixed contracture, *Gait Posture* (2004) 273–279 (England).
[84] A. Aguilera, A. Subero, Automatic gait classification patterns in spastic hemiplegia, Adv. Data Anal. Classif. 14 (2020) 897–925.
[85] A. Ferrari, L. Bergamini, G. Guerzoni, S. Calderara, N. Bicocchi, G. Vitetta, et al., Gait-Based Diplegia Classification Using LSMT Networks, J. Health Eng. 2019 (2019) 3796898.
[86] A. Krechowicz, S. Deniziak, D. Kaczmarski, Machine learning approach to gait deviation prediction based on isokinetic data acquired from biometric sensors, Gait Posture 101 (2023) 55–59.
[87] M. Sangeux, J. Rodda, H.K. Graham, Sagittal gait patterns in cerebral palsy: the plantarflexor-knee extension couple index, Gait Posture 41 (2) (2015) 586–591.
[88] E. Dolatabadi, A. Mansfield, K.K. Patterson, B. Taati, A. Mihailidis, Mixture-model clustering of pathological gait patterns, IEEE J. Biomed. Health Inform. 21 (5) (2016) 1297–1305.
[89] J. Hu, L. Zhang, J. Li, Q. Wang, A new mixed clustering-based method to analyze the gait of children with cerebral palsy, CMC-Comput. Mater. Continua 66 (2) (2021) 1551–1562.
[90] Y. Kim, T.C. Bulea, D.L. Damiano, Children with cerebral palsy have greater stride-to-stride variability of muscle synergies during gait than typically developing children: implications for motor control complexity, Neurorehabil. Neural Repair 32 (9) (2018) 834–844.
[91] K. Kaczmarczyk, A. Wit, M. Krawczyk, J. Zaborski, Gait classification in post-stroke patients using artificial neural networks, Gait Posture 30 (2) (2009) 207–210.
[92] E. Desailly, Y. Daniel, P. Sardain, P. Lacouture, Foot contact event detection using kinematic data in cerebral palsy children and normal adults gait, Gait Posture 29 (1) (2009) 76–80.
[93] A. Merlo, I. Campanini, Applications in movement and gait analysis, Surf. Electromyogr.: Physiol., Eng., Appl. (2016) 440–459.
[94] J.R. Gage, Gait Analysis in Cerebral Palsy, Oxford, Mac Keith Press, 1991.
[95] R.O. Duda, P.E. Hart, Pattern classification and scene analysis, Vol. 3, Wiley New York, 1973.
[96] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media, Inc, 2022.