DIGITAL
HEALTH

The application of artificial intelligence and custom algorithms with inertial wearable devices for gait analysis and detection of gait-altering pathologies in adults: A scoping review of literature

Ashley Cha Yin Lim[1,4] (iD), Pragadesh Natarajan[1,2,3], R Dineth Fonseka[1,2,3], Monish Maharaj[1,2,3] and Ralph J Mobbs[1,2,3]

Abstract

Background: The purpose of this scoping review was to explore the current applications of objective gait analysis using inertial measurement units, custom algorithms and artificial intelligence algorithms in detecting neurological and musculoskeletal gait altering pathologies from healthy gait patterns.

Methods: Literature searches were conducted of four electronic databases (Medline, PubMed, Embase and Web of Science) to identify studies that assessed the accuracy of these custom gait analysis models with inputs derived from wearable devices. Data was collected according to the preferred reporting items for systematic reviews and meta-analysis statement guidelines.

Results: A total of 23 eligible studies were identified for inclusion in the present review, including 10 custom algorithms articles and 13 artificial intelligence algorithms articles. Nine studies evaluated patients with Parkinson's disease of varying severity and subtypes. Support vector machine was the commonest adopted artificial intelligence algorithm model, followed by random forest and neural networks. Overall classification accuracy was promising for articles that use artificial intelligence algorithms, with nine articles achieving more than 90% accuracy.

Conclusions: Current applications of artificial intelligence algorithms are reasonably effective discrimination between pathological and non-pathological gait. Of these, machine learning algorithms demonstrate the additional capacity to handle complicated data input, when compared to other custom algorithms. Notably, there has been increasing application of machine learning algorithms for conducting gait analysis. More studies are needed with unsupervised methods and in non-clinical settings to better reflect the community and home-based usage.

Keywords

Gait-altering pathology, wearable devices, artificial intelligence, machine learning, classification models, musculoskeletal disease, neurological disease

Introduction

Gait analysis is the study of human movement involving the assessment of quantitative parameters of walking.[1–3] Chatzaki et al.[3] proposed that gait can differ between individuals due to the relative physical and psychological state of the individual at that time. Each individual has their own unique gait, however a relatively 'normative' gait should be

[1]NeuroSpine Surgery Research Group (NSURG), Australia
[2]Neuro Spine Clinic, Prince of Wales Private Hospital, Australia
[3]Faculty of Medicine, University of New South Wales (UNSW), Australia
[4]Faculty of Health and Medicine, The University of Newcastle, Australia

Corresponding author:
Ashley Cha Yin Lim, Faculty of Health and Medicine, the University of Callaghan, Newcastle, NSW 2308, Australia.
Email: ashleylcy@live.com

seen in most individuals without neuromuscular pathology.[3] Initially, gait analyses were mostly conducted in laboratory conditions with the use of multi-camera motion capture systems and force plates, in which the gold standard technologies were the optoelectronic systems.[1,3–8] In recent times, the development of wearable sensors has mitigated several issues associated with traditional gait analysis including the cost of establishing specialised laboratories and equipments, lengthy set-up and post-processing times, as well as scheduling difficulties for patients to attend gait analysis sessions.[1,9,10] Wearable sensors are lightweight and can be used outside of laboratory environments, allowing researchers and clinicians to collect indoor and outdoor movement data from the community ('free-living gait').[1,4,7–11] Individuals are instructed to wear the wearable sensors usually around their waist, on their wrist or elsewhere in contact with the body.[1,6,12] Current devices may include a range of motion sensors: gyroscopes, accelerometers, magnetometers force sensors, goniometers, inclinometers, strain gauges and more.[1,4,12] Accurate placement of devices has also been demonstrated to mimic camera motion capture systems by providing simultaneous real-time positional data.[8] In general, as a subset most inertial wearable sensors usually include integrated accelerometers detecting linear motion along reference axes, and gyroscopes which detect angular motion during ambulation by measurement of the Coriolis force.[1,12] However, wearable sensors have their limitations. One disadvantage of wearable sensors is false step detection.[13] This occurs when non-gait activity, such as swinging legs or arms while seated, gives a false impression of periodic movement in a gait cycle.[13] For wireless wearables there are also concerns regarding data losses during data transmission towards the host computers.[14] Furthermore, Peters et al.[15] mentioned the need for further reliability testing for wearables inertial devices, especially consumer-grade FitBit that are more affordable and accessible for patients to purchase compared to research-grade wearables. Another notable disadvantage that developers aim to mitigate is the issue of drifting, especially when placed near metal or magnetic fields due to the inner magnetometers.[16,17]

Data collected from multiple motion sensors require a computing system to process the complex input.[4] The output from the wearable devices can be processed through either custom algorithms or artificial intelligence (AI) models. This allows for outpatient day-to-day gait rehabilitation monitoring, providing information regarding health diagnostic and disease severity management.[1,3,4,9,12] With recent advances in computing capabilities, researchers have been looking into incorporating more AI models due to its proposed capability to analyse large datasets and identify complex patterns in gait.[18] Machine learning (ML) models are a subset of complex AI models with supervised and unsupervised learning being the two major types.[18,19] Supervised learning methods are employed when both the

inputs and outputs are known (or 'labelled'), and these include linear regression, logistic regression (LR) and decision trees.[19–21] Unsupervised learning encourages the machine to discover new associations without human interference.[19–21] However, processed outcomes from unsupervised learning can include significant amounts of variability and can be more time-consuming to manually filter for the desired data outcomes.[18] Some examples of unsupervised learning include hierarchical clustering and self-organising maps.[19,20] For gait classification, prediction and analysis, the approaches are similar: data acquisition by wearables, followed by pre-processing (noise removal, background subtraction), feature extraction, feature selection, classification and analysis of outcome.[22]

In many of the previous works on gait analysis, ML models have been applied in ambulatory gait analysis and optical motion capture systems, whether in conjunction or as separate fields of studies.[16,17] Ambulatory gait analysis examines 'normal walking bouts' in healthy adults, such as running, walking. Regression models are generally preferred to evaluate gait parameters, such as walking speed, in ambulatory gait analysis.[17,23,24] Our present study seeks to inform researchers and clinicians of pathological gait analysis in adults using wearable technologies, in which we believe is the first of its kind. There are multiple theorems behind 'normal gait', for example, the 'six determinants of gait', 'inverted pendulum model' and 'dynamic walking'.[25] In a brief summary, pathological gait occurs when an injury or pathology affects the biomechanics of gait leading to increased energy expenditure compared to 'normal gait'.[25] Some conditions that present with pathological gait include spinal cord injuries, stroke and limb loss which there is a reduction in strength and coordination.[25]

There is no general consensus of the best suited algorithm or model for gait analysis. Carcreff et al.[13] discussed that algorithms, which commonly have fixed thresholds, are best used in specific populations of study groups, namely healthy adults or elderly patients without significant gait impairment. The authors proposed that adaptive threshold algorithms can detect the inter-step variability of out-of-laboratory gait monitoring, which can outperform some ML models such as random forest (RF) and support vector machine (SVM).[13] The reasoning behind this is that 'abnormal gait' has different gait patterns that can give a false impression of peak swing and/or stance phase and abnormally slow gait.[13]

ML models are subject to overfitting when algorithm parameters are optimised to training data but are not generalised for 'real-world' test data.[20,21] To avoid overfitting, a section of data can be removed for data validation while the remaining data set is used for training using post-hoc techniques.[20,21]

This systematic review aims to inform current applications of AI and custom algorithms in the health diagnostics of gait-altering pathologies. It also aims to inform clinicians

and researchers of the most suitable and reliable ML models for diagnosing pathological gait, reporting on the accuracy and validity of current AI models. As such, the present review identifies the most common algorithms, data and applications, current issues and future possibilities.

Methods

The PRISMA statement guidelines were followed in identifying, screening and selecting studies for inclusion, and extracting data for this present review (Figure 1).

Eligibility criteria

The focus of this review was on published articles written in English and published between 1980 and February 2021, that assessed (health diagnostic) AI algorithms developed with data from inertial wearable devices. These AI algorithms were used to identify or classify pathological gait patterns (from healthy control groups). Studies involving non-inertial wearable devices such as robotic apparatus, exoskeletons or ground reaction force sensors were excluded. Reviews, conference abstracts and books were also excluded. See Figure 1 for the PRISMA flowchart and other exclusion reasons.

Search strategy

Relevant studies were identified through a systematic search for published papers in the following scientific databases from the date of inception to February 28, 2021: Medline (OvidSP), Embase (OvidSP) and Web of Science (ClarivateTM). The search 'concepts' were *objective gait analysis*, *wearable devices* and *gait-altering (neurological and musculoskeletal) pathologies* (see Table 1).

Study selection

The literature search was completed by two authors (PN and RDF). Titles and abstracts of all studies identified were screened for relevance. Studies which were clearly not relevant based on the title and abstract screen were excluded from the review. The full text of the record was reviewed if the study appeared relevance or was of uncertain relevance and a third reviewer consulted (RJM) until a consensus agreement was reached regarding inclusion/exclusion. The full text of all selected relevant records was reviewed, and eligibility was determined using the eligibility criteria defined above. The quality of each included record was assessed by two authors (PN and AL), and relevant information was extracted.

Data extraction

Data was collected regarding participant characteristics, sensor, algorithm, data methods and accuracy for each study included in our review. Participant information included number and type of participants, age and sex. Information about the type, model, sampling frequency and location on the body of each wearable sensor was also recorded. The pathology of interest, environment of the study and model features (gait variables: gait phases, spatiotemporal metrics, joint angles) were also recorded included walking speed, distance and time. Walking speed was labelled as 'self-selected' if participants were walking at a comfortable speed for them individually, but required to maintain the same speed. If participants were allowed to walk at their own speed, it was labelled as 'not controlled'.

Quality assessment

To assess the quality of included studies, two different tools were used. To evaluate the method of data acquisition from the subjects, we applied a questionnaire inspired in the Critical Appraisal Skills Program (CASP) for Diagnostic Test Studies[23] as below:

Q1: Was there a clear question or objective for the study to address?
Q2: Was there a comparison with an appropriate reference standard (gold standard)?
Q3: Did all the participants get the diagnostic test and reference standard?
Q4: Could the results of the test have been influenced by the result of the reference standard?
Q5: Is the disease status of the tested population clearly described?
Q6: Were the methods for performing the test described in sufficient detail?
Q7: Are the results presented in a way that we can work them out?
Q8: How sure are we about the results?
Q9: Can the results be applied to the population of interests?
Q10: Can the test be applied to the population of interests?
Q11: Were all outcomes important to the population considered?
Q12: Would this impact current treatment or management for your population of interests?

We scored the studies as follows. If a study completely fulfilled a criterion, we gave it a 1.0. In absent cases, the grade was 0.0, while we rated a total of 0.5 for partly fulfilled criteria.
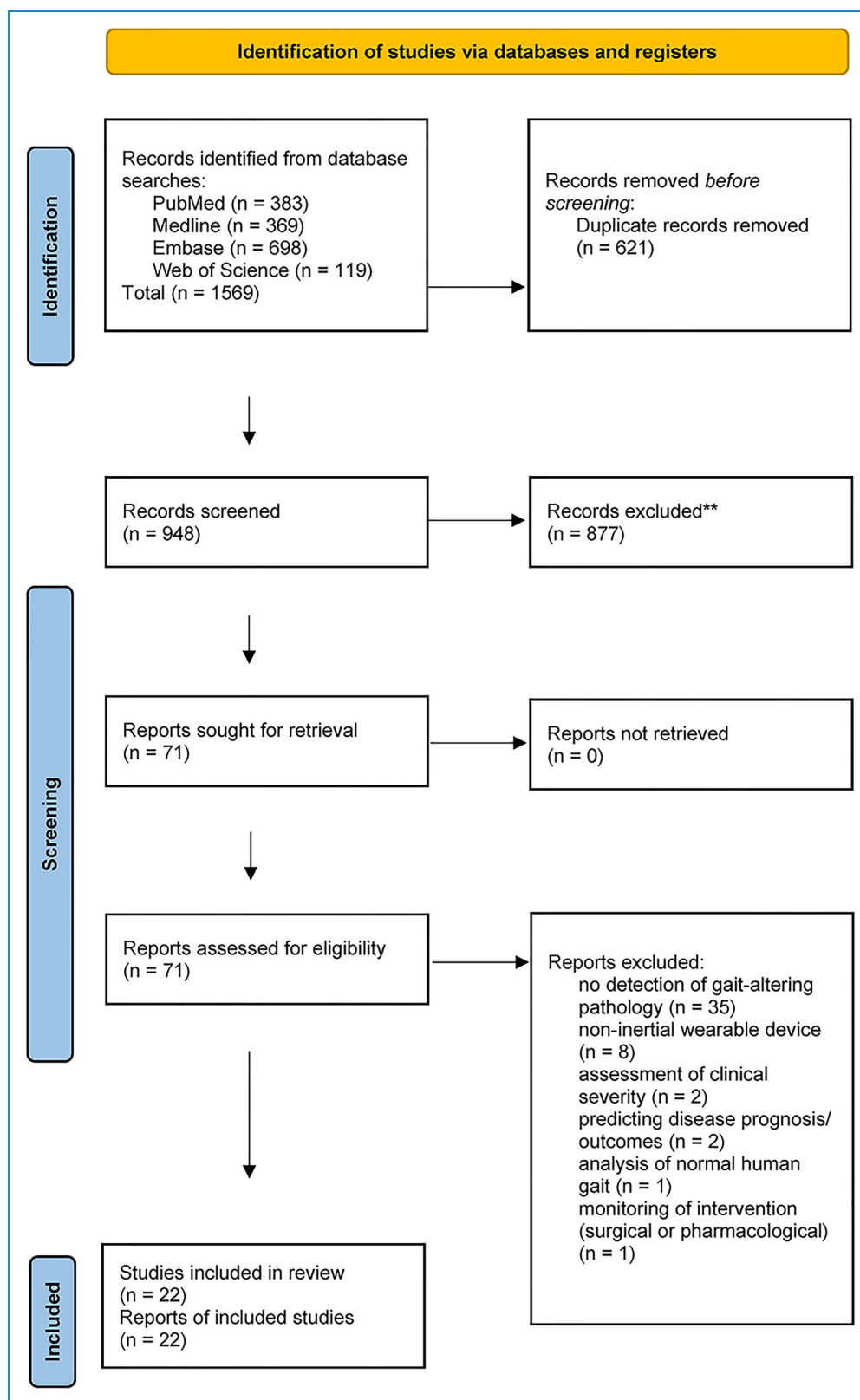
Table 1. Search strategy.

Objective gait analysis	Wearable device	Gait-altering pathologies	
1. exp Gait/(30,209) 2. exp Walking/ (56,429) 3. 1-2 OR (56,729)	4. exp Wearable Electronic Devices/ (12,883) 5. Wearable.ab,hw,kf,ti. (15,505) 6. 4 OR 5 (24,843)	Neurological: 7. exp Nervous System Disease/ (2,563,766) 8. exp Spine/(148,215) 9. exp Spinal Cord/(103,622) 10. exp Spinal Disease/(126,445) 11. exp Low Back Pain (22,620) 12. exp Brain Diseases/ (1,246,302) 13. exp Mental Disorders/ (1,271,951) 14. 7-14 OR (3,620,024)	Musculoskeletal: 15. exp Musculoskeletal Disease/ (1,104,233) 16. exp Hip/(12,088) 17. exp Knee/(14,617) 18. 15-17 OR (1,119,073) 19. 14 OR 18 (4,384,249) 20. 3 AND 6 AND 19 (374) 21. limit 22 to yr = "1980 -Current" (374) 22. limit 23 to (English language and humans) (369)

Results

Included studies

Database searches identified 1569 relevant studies. After the removal of duplicates, 948 studies remained. A total of 877 references were excluded from title and abstract screen.

A further 49 articles were excluded upon full-text review, leaving a final 21 studies to be included in qualitative synthesis (Figure 1: PRISMA flowchart). There were nine studies included in the custom algorithms category and 13 studies included in the AI category for comparison.

Study population

The most commonly investigated study population was Parkinson's disease (PD), with nine studies comparing participants with mild or unspecified severity of PD with healthy controls or another gait-altering pathology (Table 2). Other study populations included dementia, stroke, ataxia and sports injuries (post-ACL repair, concussion).

Inertial wearable sensor

Different wearable devices were employed by included studies to collect gait metrics. The number of sensors, type of sensors, device, location of devices and sampling frequency are evaluated in Table 3. The commonly used devices were Opal APDM, with L3 to L5 lower back being the preferred location for devices to be attached to participants (Table 3). The sampling frequency included in the studies varies from 50 to 160 Hz. Three studies developed custom-made wearable devices for the study: Tedesco et al.,[46] Rovini et al.[44] and Tesconi et al.[31]

Protocols

All studies included gait tasks of varying methods: Timed Up and Go (TUG),[29,30,34,35,45,47] 6MWT,[32] instrumented stand and walk task[40] and many more. The discriminating factor for TUG is that it includes the 'turning phase'. Similarly, in Di Lazzaro et al.,[35] SVM reached 97% accuracy with features extracted from Pull Test, TUG, tremor and bradykinesia items. Four studies examined participants on different speeds[37,41,44,47] while one study examined participants on maximum speed.[46] Other studies either did not specify gait speed[35] or allow participants to conduct the task at their preferred speed. Tedesco et al.[46] achieved an accuracy of 73.07% with a multilayer perceptron (MLP) model with participants conducting the task at maximum speed.

AI classification models

The most commonly used AI model was SVM, with other models including RF and neural network (NN) (see Figure 2).[40] Rehman et al.[43] used radial basis function kernel alongside SVM (SVM-RBF) in their study. One study applied back-propagation artificial NN (BP-ANN) while another study applied CNN. Both BP-ANN and CNN are classified under NN. Hsu et al.[36] and Tedesco et al.[46] applied MLP which is a class of feedforward ANN.

Discussion

Our study demonstrates that gait analysis using the AI model yields promising results in discerning pathological gait from non-pathological gait. AI technology is an emerging field of medicine in which the search for the most fitting algorithm(s) still remains. The present study,

Table 2. Study populations of included studies.

Authors	Algorithm type	Pathology	Sample size (male/female)	Mean age (years ± SD)
Buckley et al.[26]	Custom algorithm	Early-stage Parkinson disease	70 (47M/23F)	69.2 ± 9.9
		Age-matched controls	64 (35M/29F)	71.6 ± 6.8
Fino et al.[27]	Custom algorithm	Suffered a sports-related concussion SRC within 48 h	24 (18M/6F)	20.3
		Age, gender, height and mass-matched controls.	25 (19M/6F)	20.9
Ilg et al.[28]	Custom algorithm	Degenerative cerebellar ataxia (DCA)	43 (23M/20F)	51 ± 15
		Healthy controls	35	48 ± 15
Mc Ardle et al.[10]	Custom algorithm	Mild cognitive impairment/dementia due to Alzheimer disease	32	77 ± 6
		Dementia with Lewy bodies	28	76 ± 6
		Parkinson disease	14	78 ± 6
Simila et al.[29]	Custom algorithm	Older adults, independent and no cognitive incapability	35 (0M/35F)	73.8
Stack et al.[30]	Custom algorithm	Patients with and without Parkinson	24 (5M/19F) (10 with Parkinson)	74 ± 8
Tesconi et al.[31]	Custom algorithm	Vascular disease	5 (1M/4F)	73
		Healthy controls	4 (3M/1F)	32
Zhang et al.[32]	Custom algorithm	Post-stroke patients	16 (9M/7F)	54
		Healthy controls	9 (5M/4F)	35
Zhou et al.[33]	Custom algorithm	Haemodialysis patients with diabetes and end-stage renal disease	44 (25M/19F)	61.8 ± 6.7
		Cognitively-impaired patients	25 (6M/19F)	68.1 ± 8.8
De Vos et al.[34]	Artificial intelligence	Progressive supranuclear palsy	21	71
		Parkinson disease	20	66.4
		Controls	39	67.1
Di Lazzaro et al.[35]	Artificial intelligence	Parkinson disease: tremor-dominant, postural instability gait difficulties, mixed phenotype	36	63
		Controls	29	69
Hsu et al.[36]	Artificial intelligence	Stroke, brain concussion, spinal injury, brain haemorrhage	11	65.2
		Controls	9	66.4

(continued)

Table 2. Continued.

Authors	Algorithm type	Pathology	Sample size (male/female)	Mean age (years ± SD)
Jung et al.[37]	Artificial intelligence	Semi-professional athlete	29	52.92 ± 9.6
		Patients with foot deformities	21	27.9 ± 3.1
		Controls	19	29.18 ± 4.89
Kashyap et al.[38]	Artificial intelligence	Cerebellar ataxia (CA)	23 (12M/11F)	65 ± 11
		Controls	11 (6M/5F)	58 ± 12
Mannini et al.[39]	Artificial intelligence	Post-stroke patients (PS)	15 (10M/5F)	61.3 ± 13
		Huntington disease patients (HD)	17 (10M/7F)	54.3 ± 12.2
		Controls	10 (4M/6F)	69.7 ± 5.8
Moon et al.[40]	Artificial intelligence	Parkinson disease	524	66.73 ± 9.17
		Essential tremor (ET)	43	66.98 ± 9.84
Nukala et al.[41]	Artificial intelligence	Patients with balance disorders	4	N/A
		Controls	3	N/A
Rastegari[42]	Artificial intelligence	Mild Parkinson disease	10 (5M/5F)	63.8 ± 9.3
		Controls	10 (5M/5F)	64 ± 8.4
Rehman et al.[43]	Artificial intelligence	Mild Parkinson disease	93 (59M/34F)	69.2 ± 10.1
		Controls	103 (49M/54F)	72.3 ± 6.7
Rovini et al.[44]	Artificial intelligence	Idiopathic hyposmia (IH)	30 (21M/9F)	66 ± 3.2
		Parkinson disease	30 (25M/5F)	67.9 ± 8.8
		Controls	30 (25M/5F)	65.2 ± 2.5
Steinmetzer et al.[45]	Artificial intelligence	Motor dysfunction (MD)	15	N/A
		Controls	24	N/A
Tedesco et al.[46]	Artificial intelligence	Post-ACL athletes (rugby players)	6M	29.3 ± 4.5
		Controls	6M	22.8 ± 3.7

reviewed 22 papers with custom algorithms and AI algorithms.

A total of 18 of the 22 studies were published from 2018 to 2020, suggesting an exponentially increasing interest in this topic (Figure 3). There is a varied patient population selected by the papers. Nine studies included participants with PD of varying disease severity.[10,26,30,34,35,40,42–44] One study compared PD patients with progressive supranuclear palsy (PSP) which is an atypical Parkinsonian disorder that clinically shares certain features with PD.[34] This poses the potential for misdiagnosis of PD in PSP patients, therefore the use of the AI model with gait analysis has a role in supplementing the clinical diagnosis. Another study included patients with idiopathic hyposmia (IH) which is a condition that is associated with an increased risk of developing PD.[44] Other study populations

Table 3. Wearable devices employed by included artificial intelligence studies.

Authors	Device	Location of device	Number of devices	Sampling frequency
Buckley et al.[26]	OpalTM, APDM Inc, Portland, OR, USA	5th lumbar vertebra, Back of the head	2	128 Hz
Fino et al.[27]	Opal v1, APDM, Inc. Portland, OR USA	Forehead, lumbar spine, bilateral anterior distal region of shank.	4	128 Hz
Ilg et al.[28]	Opal, APDM, Inc. Portland, WA	Both feet, posterior trunk at the level of L5.	3	N/A
Mc Ardle et al.[10]	Accelerometer-based wearable (AX3, Axivity, York, UK)	L5	1	100 Hz
Simila et al.[29]	Accelerometer (GCDC X16-2)	Centre of lower back between L3-L5, anterior right hip	2	100 Hz
Stack et al.[30]	Tri-axial wearable sensors containing accelerometers and gyroscopes (2000° at 0.06 dps resolution)	Waist, ankles, wrists	5	N/A
Tesconi et al.[31]	Sensorised shoe, knee band KB and on body	Pressure sensors placed on the sole of the shoe near the heel	2	N/A
Zhang et al.[32]	3D accelerometers (MTw, Awinda, Xsens, Enschede, The Netherlands)	Midline of lower back L3 and each foot	3	100 Hz (resampled to 200 Hz using linear interpolation)
Zhou et al.[33]	Wearable sensors (LegSysTM, BioSensics LLC., MA, USA)	Lower back and dominant front lower shin for standing balance task. Sensors on both lower shin for supervised walking. Sensors worn in front of the chest for unsupervised walking	Standing balance task and gait task: 2; unsupervised gait: 1	50 Hz
De Vos et al.[34]	OpalTM, APDM, Portland, USA	Lumbar spine, right arm, right foot	6	100 Hz
Di Lazzaro et al.[35]	Movit (Captiks Srl, Rome Italy)	Hands (tremor, RAHM), thighs (LA), feet (HTT) calves and trunk (PT), arms and trunk (TUG)	9	50 Hz
Hsu et al.[36]	Delsys Trigno (Delsys Inc., Boston, MA, USA)	Lower back (L5) and both sides of the thigh, distal tibia (shank), and foot	7	148 Hz
Jung et al.[37]	Xsens MVN, Enschede, Overijssel, the Netherland	Posterior pelvis, both thighs, both shanks, both feet	7	N/A
Kashyap et al.[38]	InvenSense, Inc., San Jose, CA, USA Design Chipset "MPU-9150"	10 cm from the mouth, dorsum of hand, wrist, 1.5 m from the subject, dorsum of the foot,	1 (one device for multiple settings)	50 Hz

(continued)

Table 3. Continued.

Authors	Device	Location of device	Number of devices	Sampling frequency
		xiphisternum, midline upper back, both ankles		
Mannini et al.[39]	IMU (Opal, APDM, Inc., Portland, OR, USA) and Gait pressure mat (GAITRite TM Electronic Walkway, CIR System Inc., Franklin, NJ, USA)	Both shanks (20 mm above the malleoli), lumbar spine between L4 and S2	3	128 Hz (IMUs), 120 Hz (walkway)
Moon et al.[40]	Opal, APDM, Inc., Portland, OR, USA	On dorsal side of wrists, dorsal side of feet metatarsals, sternum, L5 lumbar	6	128 Hz
Nukala et al.[41]	Texas Instruments, Dallas, TA, USA	T4 at the back	1	160 Hz
Rastegari[42]	SHIMMER	Left and right ankle	2	102 Hz
Rehman et al.[43]	Tri-axial accelerometer (BSN Medical Limited, Hull, UK), instrumented mat (Platinum model GAITRite)	Lower back L5	1	tri-axial accelerometer (100 Hz), instrumented mat (240 Hz)
Rovini et al.[44]	SensFoot V2	Dorsum of foot	1	100 Hz
Steinmetzer et al.[45]	Meta motion rectangle wearable sensors from Mbientlab	Wrists	2	100 Hz
Tedesco et al.[46]	Custom made	(1) Anterior tibia, 10 cm below the tibial tuberosity; (2) Lateral thigh, 15 cm above the tibial tuberosity	4 (2 per leg)	100 Hz

include post-stroke patients,[32,36,39] post-concussion athletes[27,36] and dementia patients.[10]

Gait tasks with specified distance or speed such as the TUG test[30,34] were the preferred choice of most studies performing gait analysis. Some studies varied the speed and included additional movements (e.g. turning clockwise or head tilt) to examine if certain features are relevant for classification accuracy.[34,41] Kashyap et al.[38] were unique in additionally evaluating speech patterns (consistency in volume, articulation and vocal instability) by asking participants to repeat consonant-vowel syllable /ta/. Moreover, studies that evaluated PD patients such as in Di Lazzaro et al.[35] included upper limb tests: resting tremor, postural tremor and finger-nose tests. These tasks were extracted from the clinical evaluation scoring system 'Movement Disorders Society Unified Parkinson's Disease Rating Scale' (MDS-UPDRS). Several studies also used clinical tools such as MDS-UPDRS beyond determining the clinical severity of PD in study participants. Di Lazzaro et al.[35] were informative in utilising MDS-UPDRS tasks as their motor

tasks. Furthermore, Di Lazzaro et al.[35] demonstrated sensitivity in these clinical tools as there were nine features out of 19 features from Pull Test and TUG that were selected for the AI model. No subjective abnormalities were rated by clinicians in that cohort. This reinforces the role of wearable sensors technologies in being able to detect subtle movement abnormalities that are not discernible to visual observation (Tables 4 to 7).

The classification accuracy of algorithms may be affected by multiple factors ranging from gait tasks to the selection of features and algorithms. There was not a huge discrepancy between results from studies that included TUG versus results from studies that had 'normal' gait tasks. Therefore, the authors recommend future research to include TUG as an additional task, where feasible, without omitting the use of 'normal' gait task. In Najafi et al.,[7] researchers found that as the distance of the gait tasks increases, the participants' speed increases about 3–15%. A total of 16 out of 23 papers did not specify the speed or instruct the participants to carry out the task at their 'preferred speed'.[7]
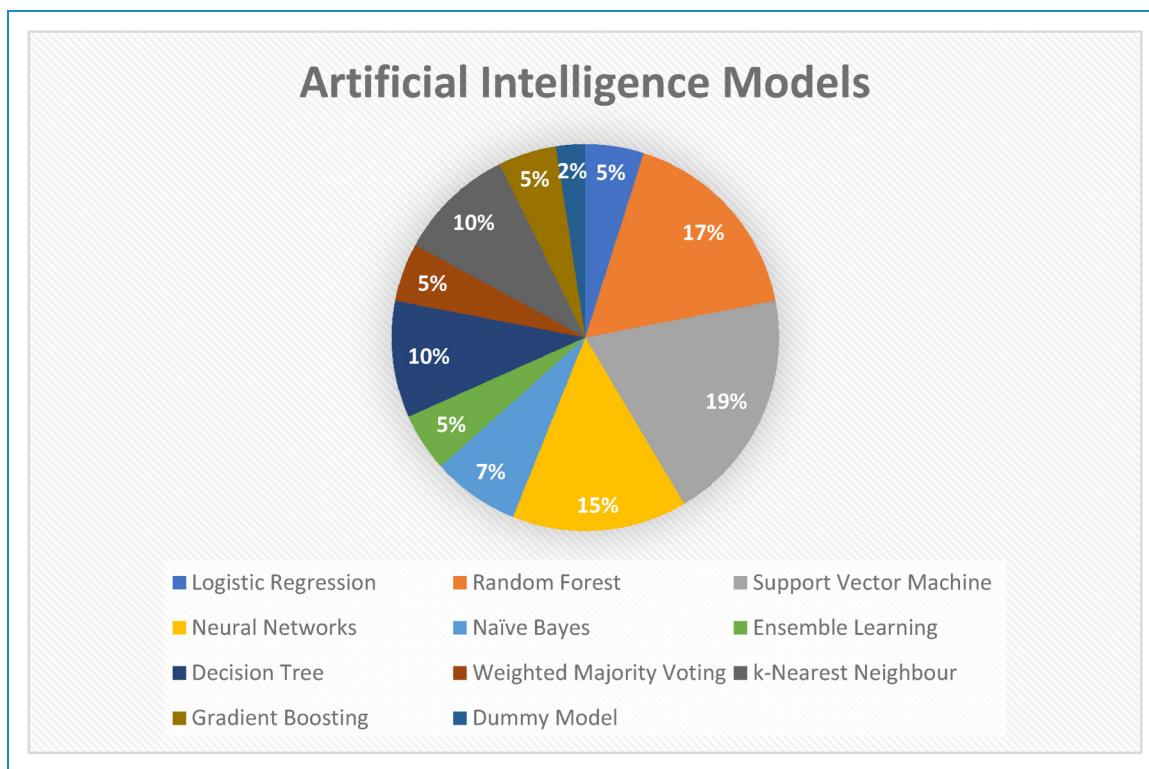
Figure 2. Artificial intelligence models employed by included studies.
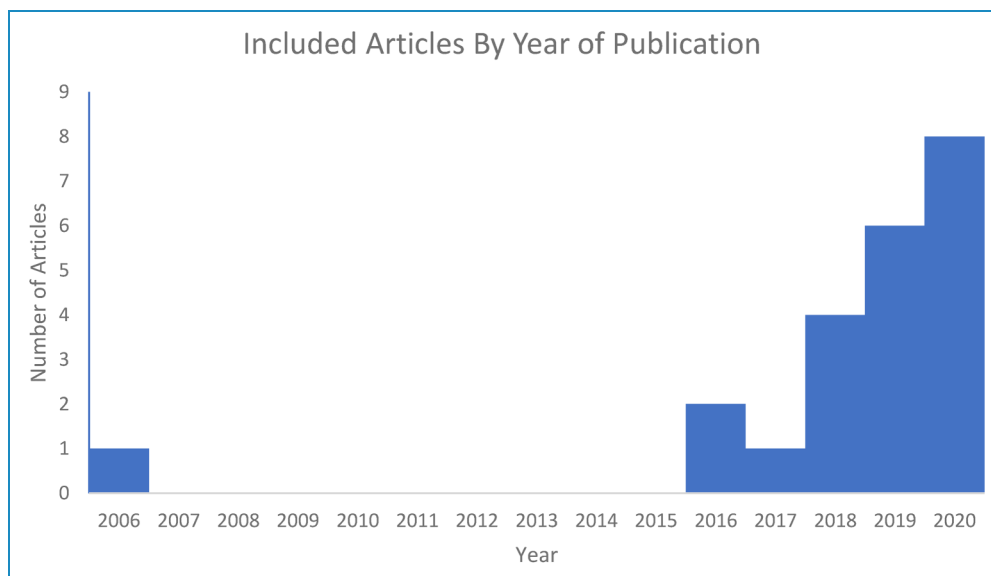

Figure 3. Increasing research interest in artificial intelligence-driven gait-analysis.

Testing environments

Studies were generally conducted in a supervised clinical environment, which is on a level surface and an obstacle-free environment with only four studies conducted the tasks in home environments or varied testing locations.[28–30,48] The results may not be suitable to extrapolate to at-home and in-community environments. This is a limitation on external validity as the patients might find gait and motor tasks easier to conduct in a clinical environment, while pathological gait might be further exaggerated in certain environments and time of the day. Various authors such as Brodie et al.[6] have previously detailed the

Table 4. Walking bout of included studies.

Authors	Task	Environment	Speed	Distance or time
Buckley et al.[26]	Gait task	Supervised clinical environment. Level surface.	Self-selected	2 min around a 25 m circuit
Fino et al.[27]	Gait task	Supervised clinical environment. Level Surface. Varied testing location.	Self-selected	2 min
Ilg et al.[28]	Gait task	(1) Supervised walking. Level surface; (2) Supervised free walking (indoors and outdoors); (3) Unsupervised walking	Self-selected	(1) 50 m; (2) 5 min; (3) 4–6 h
Mc Ardle et al.[10]	Gait task	Supervised clinical environment. Level surface	Self-selected	6×10 m
Simila et al.[29]	Berg Balance Scale BBS, Timed-Up and Go-test TUG, five times sit to stand test STS-5, 2×20 m walk	Home environment	Self-selected	2×20 m for gait task
Stack et al.[30]	(1) Chair transfer; (2) TUG; (3) Standing-start 180° turn test; (4) 3 m walk; (5) Tandem walk; (6) Rising-to-walk; reaching high and low	Supervised clinical environment and home environment	Self-selected	3 m for gait task
Tesconi et al.[31]	Gait task	Supervised clinical environment.	Self-selected	5 gait-cycle
Zhang et al.[32]	15 m to 25 m walkway with 90 degree turns (6MWT)	Supervised clinical environment. Level surface.	Self-selected	6 min
Zhou et al.[33]	(1) Standing balance task; (2) Gait task; (3) Gait task + working memory (dual-task condition); (4) Unsupervised walking	Supervised and unsupervised environment	Self-selected	15 m
De Vos et al.[34]	(1) Gait task; (2) Sway test; (3) TUG test	Straight and level surface	Self-selected	2 min
Di Lazzaro et al.[35]	Rest tremor, postural tremor, rapid alternating hand movement, foot tapping (leg agility), 20 heel-to-toe tapping, Timed-Up-and-Go test (TUG), pull test	Clinical environment	Unspecified	Unspecified
Hsu et al.[36]	First-two strides of forward direction walk	Level surface	Self-selected	6 trials of 12 m

Table 4. Continued.

Authors	Task	Environment	Speed	Distance or time
Jung et al.[37]	Gait task	Clinical environment, straight path	(1) Self-selected; (2) Fast speed; (3) Slow speed	20 m
Kashyap et al.[38]	Speech: repeated syllable utterance SPE; UL: rhythmic finger tapping test FIN, finger-nose test FNT, dysdiadochokinesia DDK, ballistic tracking BAL; LL: Heel-shin test HST, rhythmic foot tapping FOO; Balance: Rombergs ROM; Gait: gait test WAL	Clinical environment, level surface	Self-selected	10×5 m (for gait test)
Mannini et al.[39]	Gait task	Clinical environment, level surface (12 m walkway)	Self-selected	1 min
Moon et al.[40]	Instrumented stand and walk test (stand for 30 s, walk 7 m, 180° turn and walk 7 m back)	Clinical environment, level surface	Self-selected	14 m
Nukala et al.[41]	(1) Level surface; (2) Change in gait speed; (3) With horizontal head turns; (4) With vertical head turns; (5) With pivot turn; (6) Step over obstacle; (7) Step around obstacles	Clinical environment	Test 2: different gait speed; Other tests: self-selected pace	N/A
Rastegari[42]	Walk 10 m four times on an obstacle free environment, turn around after each 10 m	Clinical environment, level surface, obstacle-free	Self-selected	4×10 m
Rehman et al.[43]	(1) 4×10 m walkway (intermittent walk-IW); (2) 2 min on 25 m oval circuit (continuous walk-CW)	Clinical environment, level surface	Self-selected	(1) 4×10 m; (2) 2 min
Rovini et al.[44]	All motor tasks were performed twice. Once for R foot, once for L foot. (1) Toe-tapping for 10 s; (2) Tap their heel for 10 s; (3) Gait task; (4) Turn clockwise or anti-clockwise 360°	Clinical environment	(1) Fast speed; (2) Fast speed; (3) Self-selected; (4) Self-selected	(1) and (2) for 10 s; (3) 15 m
Steinmetzer et al.[45]	TUG	Clinical environment	Self-selected	2×10 m
Tedesco et al.[46]	Run for 5 m, sidestep left or right (for 45° angle), then run for 3 m. Repeat the test for 10 times.	Clinical environment	Maximum speed	10×8 m

variability in cadence and step time between 'free-living' community gait and laboratory-assessed gait. As such, participants may be subjected to the 'Hawthorne effect' where they are more aware of their gait, necessitating the most valid gait analysis to be conducted in community or home environments to better reflect their 'free-living gait'.[6]

Future studies may explore gait assessments in various environments, uneven surfaces and at different times of

Table 5. Characteristics of included artificial intelligence studies of custom algorithms.

Authors	Study objectives	Sensors	Algorithm or model	Variables	Model accuracy	Quality assessment score
Buckley et al.[26]	To find out whether changes in upper body motion (accelerations) during gait is a predictor of early PD.	Pressure-sensitive mat, inertial sensors	Univariate regression analysis, multivariate regression analysis	Spatiotemporal characteristics, upper body accelerations	Univariate AUC: 0.70–0.81; Multivariate AUC: 0.88–0.91	8.5
Fino et al.[27]	To examine whether horizontal head turns when seated or walking have the clinical utility for diagnosing acute concussion.	Inertial sensors	Linear mixed models (LMMs)	Gait speed, peak head angular velocity, peak head angle, response accuracy, clinical balance	Peak head angular velocity: $0.7 < AUC < 0.8$; peak head angle: $0.6 < AUC < 0.7$	8.0
Ilg et al.[28]	To identify gait features that allow quantification of ataxia-specific gait features in real life (participants with cerebellar ataxia).	Inertial sensors	Kruskal-Wallis test, Mann-Whitney U-test, Friedman test, Wilcoxon signed-rank test	Stride variability, lateral step variability	Lateral step deviation and a compound measure of spatial step variability: 0.86 accuracy	9.0
Mc Ardle et al.[10]	To differentiate dementia subtypes (AD, DLB, PDD) using gait analysis.	Inertial sensors, instrumented walkway	One-way analysis of variance (ANOVA), Kruskal-Wallis test	Pace, Variability, Rhythm, Asymmetry, Postural Control	Wearable sensors: 7 out of 14 gait characteristics; instrumented walkway: 2 out of 14 gait characteristics showed significant group differences	9.5
Simila et al.[29]	To predict early signs of balance deficits using wearable sensors.	Inertial sensors	Mann-Whitney U-test, fast Fourier transform (FFT) algorithm. Generalised linear models. Sequential forward floating selection (SFSS) method and ten-fold cross-validation	Step time, stride time	AUC 0.78 is predicting decline in total Berg Balance Scale (BBS) and 0.82 for one leg stance.	8.0
Stack et al.[30]	To evaluate the usability of wearable sensors in detecting balance impairments in people	Inertial sensors, video analyst	N/A	Stability, subtle instability (caution and near-falls),	Ratings agreed in 86/117 cases (74%) for both video analysts and wearable	7.5

Table 5. Continued.

Authors	Study objectives	Sensors	Algorithm or model	Variables	Model accuracy	Quality assessment score
	with parkinson Disease in comparision with traditional methods (observation).			time taken, parkin activity scale (PAS)	sensors data. (*highest for chair transfer, TUG, 3 m walk*)	
Tesconi et al.[31]	To investigate the possibility of using wearable sensors for monitoring flexion-extension of the knee joint during deambulation.	Knee-band, wearable sensor, and sensorised shoe	N/A	Voltage level (flexion-extension signals), irregularity parameter (gait discontinuity)	Central sensors: sensitivity 80% specificity 75%; lateral sensors: sensitivity 80% specificity 100%	7.0
Zhang et al.[32]	To differentiate post-stroke patients from healthy controls using wearable sensors and proposed gait symmetry index GSI.	Inertial sensors	Wilcoxon test; Cliff's delta; Spearman Correlation; Pearson correlation coefficient	Spatiotemporal parameters, foot pitch angular velocity	The proposed GSI of L3 has good discriminative power in differentiating post-stroke patients.	8.5
Zhou et al.[33]	To examine whether remotely monitoring mobility performance can help identify digital measures of cognitive impairments in haemodialysis patients.	Inertial sensors	Analysis of variance (ANOVA); Analysis of chi-squared, Analysis of covariance (ANCOVA), univariate and multivariate linear regression model, binary logistic regression analysis	Cumulated posture duration, daily walking performance, postural-transition	Highest AUC 0.93 model include demographics and all variables (accuracy of 85.5%)	8.5

Table 6. Characteristics of included artificial intelligence studies of machine learning models.

Authors	Study objectives	Sensors	Algorithm or model	Classification accuracy model validation	Classification accuracy results	Quality assessment score (total score 12)
De Vos et al.[34]	To investigate the use of wearable sensors coupled with machine learning as a means of disease classification between progressive supranuclear palsy and Parkinson's disease.	Inertial sensors	One-way-analysis of variance (ANOVA), independent t-test, least absolute shrinkage and selection operator LASSO, LR, RF	10-fold cross validation	93% (RF: PSP vs. HC), 88% (RF: PSP vs. PD), 93% (LR: PSP vs. HC), 80% (LR: PSP vs. PD)	10.0
Di Lazzaro et al.[35]	To assess motor performances of a population of newly diagnosed, drug-free Parkinson's disease (PD) patients using wearable inertial sensors and to compare them to healthy controls and differentiate PD subtypes (tremor dominant, postural instability gait disability and mixed phenotype).	Inertial sensors	RelieF ranking, Kruskal-Wallis feature-selection methods, SVM, one-way ANOVA	Leave-one-out cross-validation	97% (0.96)	7.5
Hsu et al.[36]	To conduct a comprehensive analysis of the placement of multiple wearable sensors for the purpose of analysing and classifying the gaits of patients with neurological disorders.	Inertial sensors	RF, neural network with multilayer perceptron, Gaussian Naïve Bayes, Adaboost, decision tree	5-fold cross validation	81% (RF), 78% (MP), 87% (NB), 84% (AB), 80% (DT)	9.0
Jung et al.[37]	To develop and evaluate neural network-based classifiers for effective categorization of athlete, normal foot, and deformed foot groups' gait assessed using wearable IMUs.	Inertial sensors	Neural network-based classifier (using gait spectrograms) CNN	4-fold cross-validation	Neural network Gait parameters: 93.02%. Gait spectrograms of foot (88.82%); post pelvis (94.25%); foot and post pelvis (96.68%); foot, shank, thigh, post pelvis (98.19%)	8.0
Kashyap et al.[38]	The study examines the potential use of a comprehensive sensor-based approach for objective evaluation of cerebellar ataxia in five domains (speech, upper limb, lower limb, gait and balance) through the instrumented versions of nine bedside neurological tests.	Inertial sensors, Kinect depth camera	RF, principal component analysis (PCA)	Leave-one-out cross-validation, comparing mean squared errors MSE by classification of various leaf size (l)	Combined nine tests demonstrate performance accuracy: 91.17%; classifying principal components PC in decreasing order of importance, performance accuracy: 97.06%; RF 97% (F1 score = 95.2%)	7.5

Table 6. Continued.

Authors	Study objectives	Sensors	Algorithm or model	Classification accuracy model validation	Classification accuracy results	Quality assessment score (total score 12)
Mannini et al.[39]	To propose and validate a general probabilistic modelling approach for the classification of different pathological gaits (healthy elderly, post-stroke patients, patients with Huntington disease).	Inertial sensors	Hidden Markov models (HMMs), SVM, majority voting (MV) classification	Leave-one-subject-out (LOSO) cross-validation	Overall accuracy: 66.7%; SVM classifier (HMM-based features only): 71.5%; SVM classifier (time and frequency domains only): 71.7%; SVM classifier (full feature): 73.3%; apply MV classifier with the full feature: 90.5%	9.5
Moon et al.[40]	To determine whether balance and gait variables obtained with wearable inertial motion sensors can be utilised to differentiate between Parkinson's disease (PD) and essential tremor (ET) using machine learning.	Inertial sensors	Neural networks, SVM, k-nearest neighbour, decision tree, RF, gradient boosting, LR, dummy model	Accuracy, recall, precision, F1 score, 3 fold cross-validation with grid search strategy	Accuracy 0.65 (kNN) to 0.89 (NN); precision 0.54 (SVM, kNN, DT, LR) to 0.61(NN); recall 0.58 (DT) to 0.63 (kNN, GB), F1 score 0.53 (DT, LR) to 0.61 (NN)	8.0
Nukala et al.[41]	To accurately classify patients with balance disorders and normal controls using wearable inertial sensors using machine learning and to determine the best-performing classification algorithm.	Inertial sensors	Back propagation artificial neural network (BP-ANN), support vector machine (SVM), k-nearest neighbours algorithm (KNN), binary decision tress (BDT)	Confusion Matrix, sensitivity, specificity, precision, negative predictive value (NPV), F-measure	BP-ANN (100%), SVM (98%), KNN (96%), BDT (94%)	7.0
Rastegari[42]	To compare the bag-of-words approach with standard epoch-based statistical approach of feature engineering methods in Parkinson disease patients.	Inertial sensors	Linear SVM, decision tree, RF, k-nearest neighbour KNN	Leave-one-subject-out cross validation LOSOCV, Pearson correlation analysis	Bag of words approach (90%), epoch-based statistical feature (60%)	7.0
Rehman et al.[43]	To compare the impact of walking protocols and gait assessment systems on the performance of a support vector machine (SVM) and random forest (RF) for classification of Parkinson's disease.	Inertial sensors	Multivariate analysis of variance (MANOVA), independent t-test, receiver operating characteristics analysis (ROC), SVM with radial basis function SVM-RBF, RF	10-fold cross-validation with area under the curve AUC	SVM performed better than RF. Intermittent walkway (IW): no differences between Axivity and GAITRite; continuous walkway (CW): classification more accurate with Axivity (AUC 87.83) versus GAITRite (AUC 80.49)	9.0

Table 6. Continued.

Authors	Study objectives	Sensors	Algorithm or model	Classification accuracy model validation	Classification accuracy results	Quality assessment score (total score 12)
Rovini et al.[44]	To accurately classify gait altering pathologies (healthy controls, Parkinson's disease, idiopathic hyposmia) by performing comparative classification analysis using three supervised machine learning algorithms.	Inertial sensors	Kolmogorov–Smirnov test, Kruskal–Wallis test, Wilcoxon rank-sum test, Spearman correlation coefficient, SVM, RF, Naïve Bayes NB	Sensitivity, specificity, precision, accuracy, F-measure, 10-fold cross-validation	RF (0.97 accuracy); SVM (0.93 accuracy), 0.97 specificity, 0.95 sensitivity, 0.95 specificity, 0.95 accuracy); NB (0.95 accuracy). SVM had the worst performance.	6.5
Steinmetzer et al.[45]	To determine whether arm swing could accurately classify patients with Parkinson's disease using wearable sensors and machine learning algorithms.	Inertial sensors	Wavelet transformation, convolutional neural network CNN single signal, multi-layer CNN, weight voting	3-fold cross-validation, binary confusion matrix	Three-channel CNN achieved the best results. classification accuracy of 93.4% using wavelet transformation and three-layer CNN architecture	9.0
Tedesco et al.[46]	To investigate the ability of a set of inertial sensors worn on the lower limbs by rugby players involved a change-of-direction (COD) activity to differentiate between healthy and post-ACL groups via the use of machine learning.	Inertial sensors	k-nearest neighbour kNN, Naïve Bayes NB, SVM, gradient boosting tree XGB, multi-layer perceptron MLP, stacking	Leave-one-out cross-validation LOSO-CV	Multilayer perception accuracy 73.07%; gradient boosting sensitivity 81.8%. Worst accuracy SVM 71.18%. The overall accuracy is uniform among all models (between 71.18% and 73.07%)	9.5

Abbreviations : LR, Logistic Regression; RF, Random Forest; SVM Support Vector Machine; ANOVA, One-way-Analysis Of Variance.

Table 7. CASP methodological quality assessment of studies of custom algorithms and machine learning models.

Authors	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Score
Buckley et al.[26]	1.0	1.0	1.0	0.5	0.5	0.5	1.0	1.0	0.5	0.5	0.5	0.5	8.5
Fino et al.[27]	1.0	0.0	0.5	1.0	0.5	0.5	0.5	1.0	0.5	1.0	1.0	0.5	8.0
Ilg et al.[28]	1.0	0.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5	0.5	9.0
Mc Ardle et al.[10]	1.0	1.0	1.0	0.5	1.0	0.5	1.0	1.0	1.0	0.5	0.5	0.5	9.5
Simila et al.[29]	1.0	0.0	0.5	1.0	1.0	0.5	1.0	1.0	0.5	0.5	0.5	0.5	8.0
Stack et al.[30]	1.0	1.0	1.0	0.5	0.0	0.5	0.5	0.5	1.0	0.5	0.5	0.5	7.5
Tesconi et al.[31]	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.5	1.0	0.5	0.5	0.5	7.0
Zhang et al.[32]	1.0	0.0	0.5	1.0	0.0	1.0	1.0	0.5	1.0	1.0	1.0	0.5	8.5
Zhou et al.[33]	1.0	0.0	0.5	1.0	1.0	0.5	1.0	1.0	1.0	0.5	0.5	0.5	8.5
De Vos et al.[34]	1.0	0.0	0.5	1.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	1.0	10.0
Di Lazzaro et al.[35]	1.0	0.0	0.5	1.0	1.0	1.0	1.0	0.5	0.5	0.5	0.5	0.0	7.5
Hsu et al.[36]	1.0	0.0	0.5	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	9.0
Jung et al.[37]	1.0	0.0	0.5	1.0	0.0	1.0	1.0	1.0	1.0	0.5	1.0	0.0	8.0
Kashyap et al.[38]	1.0	1.0	1.0	0.5	0.0	1.0	0.5	0.5	0.5	0.5	0.5	0.5	7.5
Mannini et al.[39]	1.0	1.0	1.0	0.5	0.5	0.5	1.0	1.0	1.0	1.0	0.5	0.5	9.5
Moon et al.[40]	1.0	0.0	0.5	1.0	0.0	1.0	0.5	1.0	0.5	1.0	1.0	0.5	8.0
Nukala et al.[41]	1.0	0.0	0.5	0.5	0.0	1.0	1.0	0.5	0.5	1.0	1.0	0.0	7.0
Rastegari[42]	1.0	0.0	0.5	1.0	0.0	0.5	1.0	0.5	0.5	1.0	0.5	0.5	7.0
Rehman et al.[43]	1.0	1.0	1.0	0.5	0.5	1.0	0.5	1.0	0.5	0.5	1.0	0.5	9.0
Rovini et al.[44]	1.0	0.0	0.5	1.0	0.0	0.5	0.5	0.5	0.5	1.0	0.5	0.5	6.5
Steinmetzer et al.[45]	1.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.5	1.0	0.5	1.0	9.0
Tedesco et al.[46]	1.0	0.0	0.5	1.0	1.0	0.5	1.0	1.0	0.5	1.0	1.0	1.0	9.5

the day. Most studies included were cross-sectional, hence a longer follow-up period to monitor changes in gait following disease severity and recovery may be warranted. The follow-up period of gait analysis could provide conformational findings such as in Rovini et al.[44] which hypothesised patients who were classified as IH have a higher likelihood of developing PD. As another example exemplifying the utility of follow-up, the Fino et al.[27] study of athlete post-acute concussion had a short-term follow-up of 10 days for all participants. Peak head angular velocity

which is the main discriminating factor found in the study in classifying athletes was found to return to healthy control levels.

Machine learning

For AI technology there are various advantages and disadvantages of using one model from another. The selection of features and algorithms may play a major role in model performance when classifying pathological gait from healthy controls.

Feature selection methods allow for training datasets to be more precise for models and reduce the training times by removing redundant features. Zhang et al.[32] applied the Pearson correlation coefficient which is one of the measures of the filter method. Other types of statistical tools used within the filter method include the chi-squared and analysis of variance (ANOVA) test. In Rehman et al.,[43] Di Lazzaro et al.[35] and De Vos et al.[34] the statistical tool 'analysis of variance' were used. Di Lazzaro et al.[35] used Kruskal-Wallis feature selection which is a non-parametric ANOVA test of the wrapper method and also relief ranking which is a type of filter method. The wrapper method is more complex than the filter method and therefore would perform comparatively better with a smaller dataset. It is also more computationally intensive than the filter method and can be subject to overfitting as model complexity increases. A benefit of the filter method is that it does not incorporate a specific ML algorithm. The embedded method combines features from both filter and wrapper methods. The best example of an embedded method is LASSO, or 'least absolute shrinkage and selection operator'. De Vos et al.[34] implemented both ANOVA and LASSO feature selection methods. Similar to the wrapper method, the embedded method also has a high computational requirement.

Feature selection and extraction

An important concept for feature selection is 'bias-variance tradeoff'.[49] Bias reflects how much model predictions differ from the correct value and variance is how much the predictions for a given point vary between different realizations of the model.[49] The ideal scenario for feature selection is when the authors achieve a balance between underfitting (high bias) and overfitting (high variance).[49] This can be prevented with model validation tools such as leave-one-out cross-validation and n-fold cross-validation. The bias–variance relationship can be explained along with out-of-sample error and in-of-sample error.[49] In-of-sample error will drop initially with the increase in data, however as the amount of data increases the error increases.[49] Out-of-sample error follows a different trend. It will decrease as the number of data increases.[49] As we approach infinite data, the in-of-sample error and out-of-sample error will reach the same value.[49] This is known as the 'bias' of the model. Five studies applied leave-one-out cross-validation model;[35,38,39,42,46] Three studies applied 10-fold cross-validation model.[34,43,44] It is difficult to determine which method is the most suitable validation tool, as the results can be subjected to multiple confounding variables in the studies. However, leave-one-out cross-validation and n-fold cross-validation (with a higher number of n values) would be the recommended validation tools by the authors as it is more rigorous in its implementation.

Feature extraction differs from feature selection in that new features are created from existing features. This method minimises the number of features, by discarding the original features and improves the performance of the model. Principle components analysis (PCA) is one of the most used linear dimensionality reduction method, as in Kashyap et al.[38]

SVM, RF and LR

In Hsu et al.,[36] Moon et al.[40] and Tedesco et al.,[46] ensemble methods were implemented which can help in preventing overfitting. Classically, studies favour the use of hidden Markov model (HMM)-based features and SVM classifier, as highlighted by Hsu et al.[36] This is further supported by eight out of 13 studies in ML which applied the SVM algorithm. SVM is a type of supervised learning and can process large datasets with good generalisation capabilities. It generally has a good predictive accuracy as shown in Nukala et al.[41] of 98% accuracy. However, three out of eight studies found that SVM was not the best classifier in terms of application. In Nukala et al.,[41] BP-ANN yielded a 100% classification accuracy. Rovini et al.[44] compared SVM with RF and NB, in which SVM and NB achieved 95% accuracy meanwhile RF yield 97% accuracy. In Tedesco et al.,[46] SVM classifiers showed the performance accuracy of only 71.18%. One of the benefits of the SVM classifier is that it is based on maximum margin hyperplane, and support vectors are the closest elements to the decision surface, therefore it avoids data overfitting.[21] In Rehman et al.,[43] RBF-SVM was used where RBF is a nonparametric algorithm and it functions to convert non-linear problems into linear, therefore simplifying the SVM model. Following that the results showed that RBF-SVM performed better than RF.[43] RF algorithm is a supervised learning model that is derived from multiple decision trees.[19] It is also a class of ensemble learning method, however, was not categorised together with (gradient) bagging and boosting in the pie chart (Figure 2).[49] RF is widely used in various gait analyses and has shown promising results in Rovini et al.[44] The benefits of this algorithm is that it is robust in handling outliers, can work with large datasets and requires simple parametrization. This is due to the random nature of partitioning in which the features are selected for splitting.

One of the other classical AI algorithms is LR employed in two out of 13 included studies.[34,40] LR uses a sigmoidal curve to explore the relationship between features and the probability of an outcome.[19]

Neural networks

ANN and CNN are classes of NNs but have different functionality. Each ANN has nodes (or neurons) that take in inputs, and synapses with other nodes via connections.[19]

The algorithm separates the classes by a line, plane or hyperplane in a two- or multi-dimensional space.[19] This is similar to SVM in which classes are divided by hyperplanes. Furthermore, features are transformed using the sigmoid function providing class associations as an output, not probability.[19] It has a high tolerance to noisy data and is suitable for untrained continuous data.[19] The robustness of NN performance is shown when processing complex data input from an unsupervised dynamic environment.[40] CNN is widely used in image classification and segmentation tasks.[19] It passes patches of an image from one node to another, the nodes would then apply convolutional filters to extract specific features of the image. Multiple layers can be conducted in which new filters are used each time and then feedforwarded into ANN. Steinmetzer et al.[45] adopted single-layer and multi-layer CNN for gait analysis. The final results show a promising result of 93.4% accuracy using wavelet transformation and a three-layer CNN algorithm.[45] Despite the complexity of NN mapping multi-dimensional images, the computational requirements of CNN are attainable from consumer-based products. For example, Steinmetzer et al.[45] applied a three-layer model CNN with consumer-grade Intel Core i7-6700HQ 2.6 GHz four cores and 16 GB RAM. The training of the CNN model took around 45 min to complete.[45] Deep learning models remove the need for manual feature selection process, conducted by experienced researchers or data scientists. In general SVM and RF, albeit commonly used, show varying results in terms of classification accuracy. Although NNs are used in a limited number of studies, the results are promising and should be further explored in future gait analysis studies.

Limitations

One of the study's main limitations is a lack of consistent measures for assessing model accuracy. Another limitation is this study did not take into account software variety and whether it would affect the performance of algorithms. In this study, the authors documented the location of wearable sensors and noted the most preferred location by researchers, however, no conclusions regarding the ideal number and location of wearable sensors could be made. There is a possibility that increasing the number of sensors used may be detrimental to the accuracy of the results, however, this is yet to be definitively determined. This review did not include paediatric gait analysis studies, therefore the findings are not generalisable to the paediatric population.

For the AI algorithms studies, there are variations in the application of AI algorithms. All studies applied AI algorithms in their 'walking segment', however, not all studies included the 'turning phase'. This may affect the classification accuracy especially for the TUG gait tasks where the 'sit-to-stand', 'stand-to-sit' and 'turning' phase

provides valuable information in discerning pathological gait. This is seen in the study conducted by Steinmetzer et al.[45] where the completed TUG task achieved an accuracy of 93.3% compared to gait alone of 90.3%. For non-walking segments, AI algorithms were applied in tasks, such as speech, balance, upper and lower limb movement tasks.

Feasibility

Unlike deep learning models, the conventional ML model requires less data input. For example within the study conducted by Hsu et al., the number of hidden layers used for NN with MLP was 2000, meanwhile the RF classifier used 1000 trees. The study analysed the first two strides of six successful gait trials. In the study conducted by Terrier,[50] the minimum number of strides was two continuous strides to reconfigure a pretrained CNN model to identify previously unseen gait. However, this is completed under a supervised laboratory environment with pressure-sensitive mat. When deciding on the size of the training set, there are no specific reference values for each ML model. The recommended rule is to plot a learning curve (LC). Model-set-curve is plotted with training-set-curve to determine whether there is a likelihood of high bias or variance.

In summary, the authors recommend the use of NNs, in particular CNN and BP-ANN. Even though it has high computational requirements, it delivered consistent accuracy (above 89%) across multiple studies.[33] Moreover, it incorporates characteristics of SVM which is the most commonly adopted AI algorithm, as evaluated in this study.

Conclusions

The results of the present review show the need for more research in the field of paediatric gait pathologies, types of Parkinsonism and analysis of free-living gait in the community and at home. It is difficult to determine which algorithms for feature selection, feature reduction, classification and validation are the most suitable for gait analysis. Therefore, there is a need for larger cohort studies to be conducted with various algorithms for direct comparison of classification accuracy. Duration of follow-up should be considered. The next step would also include longitudinal gait monitoring of patients on trials of disease-modifying medications on or off medications. The results from these can be used for prediction analyses of disease progression. Furthermore, considerations regarding the bias-variance tradeoff and the technological requirements are warranted in future development and uses of AI algorithms for gait analysis.

ORCID iD: Ashley Cha Yin Lim 🆔 https://orcid.org/0000-0002-1370-1506

References

1. Tao W, Liu T, Zheng R, et al. Gait analysis using wearable sensors. *Sensors (Basel)* 2012; 12: 2255–2283.
2. Cappozzo A. Gait analysis methodology. *Hum Mov Sci* 1984; 3: 27–50.
3. Chatzaki C, Skaramagkas V, Tachos N, et al. The smart-insole dataset: gait analysis using wearable sensors with a focus on elderly and Parkinson's patients. *Sensors (Basel)* 2021; 21: 2821.
4. Caldas R, Mundt M, Potthast W, et al. A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. *Gait Posture* 2017; 57: 204–210.
5. Benson LC, Clermont CA, Bošnjak E, et al. The use of wearable devices for walking and running gait analysis outside of the lab: a systematic review. *Gait Posture* 2018; 63: 124–138.
6. Brodie MA, Coppens MJ, Lord SR, et al. Wearable pendant device monitoring using new wavelet-based methods shows daily life and laboratory gaits are different. *Med Biol Eng Comput* 2016; 54: 663–674.
7. Najafi B, Khan T and Wrobel J. Laboratory in a box: wearable sensors and its advantages for gait analysis. *Annu Int Conf IEEE Eng Med Biol Soc* 2011; 2011: 6507–6510.
8. Takeda R, Tadano S, Todoh M, et al. Gait analysis using gravitational acceleration measured by wearable sensors. *J Biomech* 2009; 42: 223–233.
9. Chen S, Lach J, Lo B, et al. Toward pervasive gait analysis with wearable sensors: a systematic review. *IEEE J Biomed Health Inform* 2016; 20: 1521–1537.
10. Mc Ardle R, Del Din S, Galna B, et al. Differentiating dementia disease subtypes with gait analysis: feasibility of wearable sensors? *Gait Posture* 2020; 76: 372–376.
11. Kobsar D, Osis ST, Phinyomark A, et al. Reliability of gait analysis using wearable sensors in patients with knee osteoarthritis. *J Biomech* 2016; 49: 3977–3982.
12. Tarniță D. Wearable sensors used for human gait analysis. *Rom J Morphol Embryol* 2016; 57: 373–382.
13. Carcreff L, Paraschiv-Ionescu A, Gerber CN, et al. A personalized approach to improve walking detection in real-life settings: application to children with cerebral palsy. *Sensors (Basel)* 2019; 19: 5316.
14. Davidson P, Virekunnas H, Sharma D, et al. Continuous analysis of running mechanics by means of an integrated INS/GPS device. *Sensors (Basel)* 2019; 19: 1480.
15. Peters DM, O'Brien ES, Kamrud KE, et al. Utilization of wearable technology to assess gait and mobility post-stroke: a systematic review. *J Neuroeng Rehabil* 2021; 18: 67.
16. Ross GB, Dowling B, Troje NF, et al. Classifying elite from novice athletes using simulated wearable sensor data. *Front Bioeng Biotechnol* 2020; 8: 814.
17. Zhang H, Guo Y and Zanotto D. Accurate ambulatory gait analysis in walking and running using machine learning models. *IEEE Trans Neural Syst Rehabil Eng* 2020; 28: 191–202.
18. Khera P and Kumar N. Role of machine learning in gait analysis: a review. *J Med Eng Technol* 2020; 44: 441–467.
19. Choi RY, Coyner AS, Kalpathy-Cramer J, et al. Introduction to machine learning, neural networks, and deep learning. *Transl Vis Sci Technol* 2020; 9: 14.
20. Stafford I, Kellermann M, Mossotto E, et al. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *NPJ Digital Med* 2020; 3: 1–11.
21. Sidey-Gibbons JAM and Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* 2019; 19: 64.
22. Nithyakani P and Ukrit MF. The systematic review on gait analysis: trends and developments. *Eur J Mol Clin Med* 2020; 7: 1636–1654.
23. Zihajehzadeh S and Park EJ. Experimental evaluation of regression model-based walking speed estimation using lower body-mounted IMU. *Annu Int Conf IEEE Eng Med Biol Soc* 2016; 2016: 243–246.
24. Santhiranayagam BK, Lai D, Shilton A, et al. Regression models for estimating gait parameters using inertial sensors. In: *2011 seventh international conference on intelligent sensors, sensor networks and information processing*, Adelaide, SA, Australia: IEEE; 2011.
25. Kuo AD and Donelan JM. Dynamic principles of gait and their clinical implications. *Phys Ther* 2010; 90: 157–174.
26. Buckley C, Galna B, Rochester L, et al. Upper body accelerations as a biomarker of gait impairment in the early stages of Parkinson's disease. *Gait Posture* 2019; 71: 289–295.
27. Fino PC, Wilhelm J, Parrington L, et al. Inertial sensors reveal subtle motor deficits when walking with horizontal head turns after concussion. *J Head Trauma Rehabil* 2019; 34: E74–E81.
28. Ilg W, Seemann J, Giese M, et al. Real-life gait assessment in degenerative cerebellar ataxia: toward ecologically valid biomarkers. *Neurology* 2020; 95: e1199–ee210.
29. Similä H, Immonen M and Ermes M. Accelerometry-based assessment and detection of early signs of balance deficits. *Comput Biol Med* 2017; 85: 25–32.
30. Stack E, Agarwal V, King R, et al. Identifying balance impairments in people with Parkinson's disease using video and wearable sensors. *Gait Posture* 2018; 62: 321–326.
31. Tesconi M, Pasquale Scilingo E, Barba P, et al. Wearable kinaesthetic system for joint knee flexion-extension monitoring in gait analysis. *Conf Proc IEEE Eng Med Biol Soc* 2006; 2006: 1497–1500.
32. Zhang W, Smuck M, Legault C, et al. Gait symmetry assessment with a low back 3D accelerometer in post-stroke patients. *Sensors (Basel)* 2018; 18: 3322.
33. Zhou H, Al-Ali F, Wang C, et al. Harnessing digital health to objectively assess cognitive impairment in people undergoing hemodialysis process: the impact of cognitive impairment on

mobility performance measured by wearables. *PLoS One* 2020; 15: e0225358.

34. De Vos M, Prince J, Buchanan T, et al. Discriminating progressive supranuclear palsy from Parkinson's disease using wearable technology and machine learning. *Gait Posture* 2020; 77: 257–263.

35. Di Lazzaro G, Ricci M, Al-Wardat M, et al. Technology-based objective measures detect subclinical axial signs in untreated, de novo Parkinson's disease. *J Parkinsons Dis* 2020; 10: 113–122.

36. Hsu WC, Sugiarto T, Lin YJ, et al. Multiple-wearable-sensor-based gait classification and analysis in patients with neurological disorders. *Sensors (Basel)* 2018; 18: 3397.

37. Jung D, Nguyen MD, Han J, et al. Deep neural network-based gait classification using wearable inertial sensor data. *Annu Int Conf IEEE Eng Med Biol Soc* 2019; 2019: 3624–3628.

38. Kashyap B, Phan D, Pathirana PN, et al. A sensor-based comprehensive objective assessment of motor symptoms in cerebellar ataxia. *Annu Int Conf IEEE Eng Med Biol Soc* 2020; 2020: 816–819.

39. Mannini A, Trojaniello D, Cereatti A, et al. A machine learning framework for gait classification using inertial sensors: application to elderly, post-stroke and Huntington's disease patients. *Sensors (Basel)* 2016; 16: 134.

40. Moon S, Song HJ, Sharma VD, et al. Classification of Parkinson's disease and essential tremor based on balance and gait characteristics from wearable motion sensors via machine learning techniques: a data-driven approach. *J Neuroeng Rehabil* 2020; 17: 125.

41. Nukala BT, Nakano T, Rodriguez A, et al. Real-time classification of patients with balance disorders vs. normal subjects using a low-cost small wireless wearable gait sensor. *Biosensors (Basel)* 2016; 6: 58.

42. Rastegari E. A bag-of-words feature engineering approach for assessing health conditions using accelerometer data. *Smart Health* 2020; 16: 100116.

43. Rehman RZU, Del Din S, Shi JQ, et al. Comparison of walking protocols and gait assessment systems for machine learning-based classification of Parkinson's disease. *Sensors (Basel)* 2019; 19: 5363.

44. Rovini E, Maremmani C, Moschetti A, et al. Comparative motor pre-clinical assessment in Parkinson's disease using supervised machine learning approaches. *Ann Biomed Eng* 2018; 46: 2057–2068.

45. Steinmetzer T, Maasch M, Bönninger I, et al. Analysis and classification of motor dysfunctions in arm swing in Parkinson's disease. *Electronics (Basel)* 2019; 8: 1471.

46. Tedesco S, Crowe C, Ryan A, et al. Motion sensors-based machine learning approach for the identification of anterior cruciate ligament gait patterns in on-the-field activities in rugby players. *Sensors (Basel)* 2020; 20: 3029.

47. Sivarajah L, Kane KJ, Lanovaz J, et al. The feasibility and validity of body-worn sensors to supplement timed walking tests for children with neurological conditions. *Phys Occup Ther Pediatr* 2018; 38: 280–290.

48. Zhou Z, Wu TC, Wang B, et al. Machine learning methods in psychiatry: a brief introduction. *Gen Psychiatr* 2020; 33: e100171.

49. Mehta P, Wang CH, Day AGR, et al. A high-bias, low-variance introduction to machine learning for physicists. *Phys Rep* 2019; 810: 1–124.

50. Terrier P. Gait recognition via deep learning of the center-of-pressure trajectory. *Appl Sci* 2020; 10: 774.

51. CASP Diagnostic Study Checklist. *Critical Appraisal Skills Programme* 2019, Oxford.