Gait & Posture 120 (2025) 150–160

Contents lists available at ScienceDirect

Gait & Posture

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

Review

Reliability and validity of current computer vision based motion capture systems in gait analysis: A systematic review

Xingye Cheng [a], Yiran Jiao [a], Rebecca M. Meiring [a,b], Bo Sheng [c], Yanxin Zhang [a,*]

[a] Department of Exercise Sciences, Faculty of Science, University of Auckland, Auckland 1023, New Zealand
[b] School of Physiology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2050, South Africa
[c] School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China

ARTICLE INFO

Keywords:
Pose estimation algorithm
Gait analysis
Computer vision
Reliability
Validity

ABSTRACT

Background: Traditional instrumented gait analysis (IGA) objectively quantifies gait deviations, but its clinical use is hindered by high cost, lab environment, and complex protocols. Pose estimation algorithm (PEA)-based gait analysis, which infers joint positions from videos, offers an accessible method to detect gait abnormalities and tailor rehabilitation strategies. However, its reliability and validity in gait analysis and algorithmic factors affecting accuracy have not been reviewed.

Research question: This systematic review aims to evaluate the accuracy of PEA-based gait analysis systems and to identify the algorithmic factors impacting their accuracy.

Method: A total of 644 articles were initially identified through Scopus, PubMed, and IEEE, with 20 meeting the inclusion and exclusion criteria. Reliability, validity, and algorithmic parameters were extracted for detailed review.

Results and significance: Most included articles focus on validity against the gold standard, while limited evidence makes it challenging to determine reliability. OpenCap demonstrated an MAE of 4.1° for 3D joint angles, but higher errors in rotational angles require further validation. OpenPose demonstrated ICCs of 0.89–0.994 for spatiotemporal parameters and MAE < 5.2° for 2D hip and knee joint angles in the sagittal plane (ICCs = 0.67–0.92, CCCs = 0.83–0.979), but ankle kinematics exhibited poor accuracy (ICCs = 0.37–0.57, MAEs = 3.1°-9.77°, CCCs = 0.51–0.936). PEA accuracy depends on camera settings, backbone architecture, and training datasets. This study reviews the accuracy of PEA-based gait analysis systems, supporting future research in gait-related clinical applications of PEA.

1. Introduction

Gait analysis is widely used in a clinical setting to assist therapists with the identification of gait abnormalities, understand the mechanisms, and guide therapy [1]. Instrumented gait analysis (IGA) has long been considered the gold standard for gait assessment in research practice, recognized for its ability to deliver accurate and precise quantitative measurements of gait patterns and characteristics [2]. IGA refers to the use of a force plate, wearable sensors, and infrared cameras to capture and analyse a variety of human gait parameters (i.e., spatiotemporal, kinematic, and kinetic measures) [3]. IGA therefore provides accurate quantitative measurements of gait patterns, and research on the clinical applicability and efficacy of IGA suggests it may also enhance diagnosis, outcome prediction, and rehabilitation of various gait impairments [4]. However, IGA continues to be underutilized in clinical practice, largely due to the high costs, technical equipment, and the intricate protocols and data analysis required to collect and interpret the data. Furthermore, its consistent adoption is hindered by these technical challenges in environments where training, experience, and preferences differ among and within clinical teams [3].

Computer vision-based pose estimation algorithm (PEA), a technology that automatically detects and tracks joint positions in videos, has rapidly developed in recent years, offering a potentially alternative, feasible, and efficient approach to traditional gait analysis. This technology uses Convolutional Neural Networks (CNN) to identify, track, and extract the key joints of the human body in a two-dimensional space from the video footage recorded by Red Green Blue (RGB) cameras [5], such as OpenPose [6]. By additionally integrating strategies such as

* Correspondence to: Building 907 - Bldg. 907, Level 2, Room 235, 368 Khyber Pass, Newmarket, Auckland 1023, New Zealand
E-mail address: yanxin.zhang@auckland.ac.nz (Y. Zhang).

https://doi.org/10.1016/j.gaitpost.2025.04.016
Received 4 October 2024; Received in revised form 6 March 2025; Accepted 15 April 2025
Available online 17 April 2025

synchronization of multiple video cameras [7] or using three-dimensional (3D) heatmaps [8], 3D joint position data may also be estimated from the 2D capture. PEA-based gait analysis systems inherently address the limitations of traditional IGA labs due to their lower setup and maintenance costs [9]. Despite some limitations of PEA-based gait analysis, such as occlusion issues, predominantly sagittal plane analysis, and environmental factors, these systems still hold immense potential. With further advancements, they may become widely applicable in free-living environments, such as clinics, homes, and sports arenas, for remote clinical and rehabilitative applications.

In clinical gait analysis, the technical accuracy of a motion capture system, encompassing both validity and reliability, represents the foundational level of efficacy [10]. Since clinical gait analysis aims to differentiate between normal and abnormal gait patterns and monitor changes in walking over time [11], ensuring reliability and validity helps healthcare providers avoid over-interpreting small differences as meaningful [12] and ensures greater confidence that observed treatment effects exceed measurement error [13]. Although a growing body of literature has individually explored various aspects of the PEA-based gait analysis in clinical use, such as accuracy in assessing gait biomechanical outcomes across different populations [14–16], and its application in clinical settings [17–19], the performance (i.e., reliability and validity) of PEA-based gait analysis has not yet been systematically summarized. Moreover, no study has thoroughly examined the algorithmic factors influencing the effectiveness of PAE-based gait analysis systems. As a result, there is still insufficient evidence for patients and healthcare providers to make fully informed decisions and reach a consensus on the appropriate clinical application of the PEA-based gait analysis systems.

The aims of this systematic review are: (1) to evaluate the reliability and validity of current computer vision-based PEAs in gait analysis; and (2) to summarize the algorithmic factors that affect the accuracy in gait analysis.

2. Methods

2.1. Systematic search strategy

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline was followed when conducting the systematic review [20]. Articles written in English were searched in PubMed, Scopus, and IEEE databases in June 2024, and the published year was limited from 2012 to the present. Search terms were as follows: ((accuracy) OR (reliability) OR (validity)) AND (("gait analysis") OR ("gait assessment")) AND ((markerless motion capture) OR (portable) OR (mobile) OR (video-based) OR (vision-based) OR (pose estimation algorithm)) NOT (IMU) NOT (accelerometer) NOT (gyroscope) NOT (insole) NOT (insole). To search relevant articles as comprehensively as possible, cross-referencing was also performed.

2.2. Study selection

The initial search was conducted by the primary reviewer (XC) across three electronic databases. The results were then downloaded into the Zotero library for screening and review. Two reviewers (XC and YJ) then performed a full paper assessment based on the inclusion and exclusion criteria outlined in Table 1, using the titles and abstracts for initial screening. Notably, the articles using more than two cameras and the depth camera were also excluded because this setup requires higher hardware demands and reliance on controlled laboratory environments, significantly increasing the complexity of the setup and reducing their practical accessibility. The records from both reviewers were imported into Zotero, and the full texts of the remaining studies were downloaded. Two reviewers (XC and YJ) independently read the full texts of these studies. Subsequently, they had a face-to-face discussion to decide on the studies to be included. Any discrepancies between them were resolved through discussion, with the involvement of a third reviewer (YZ) to facilitate consensus.

Table 1
Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
The assessment system based on RGB camera-based motion capture system using computer vision-based pose estimation algorithm	Systematic or literature reviews
The assessment system used provides gait spatiotemporal and/or kinematic parameters	Book and other non-peer-reviewed publications
Reliability or validity test of gait assessment system	Conference Paper
Written in English	Number of cameras used in the system ≥ 3
	Not compared with the gold standard
	Not gait assessment system
	Not using computer vision-based pose estimation algorithm

2.3. Data extraction

After the final set of studies was selected, data extraction was carried out by the first author (XC) and verified by a second author (YJ). In cases of disagreement between the two authors, a third author (YZ) acted as a tie-breaker to resolve the conflicts. Information about the reliability and validity of computer vision-based PEA on gait analysis was extracted to answer question 2. As the searched articles rarely discussed algorithmic factors, technical documentation for each algorithm was consulted. Algorithm specifications such as keypoints (the specific joints or body parts the algorithm tracks), backbone architecture (the deep learning model responsible for feature extraction), and inference time (the time taken by the algorithm to process an image and produce results) were extracted to address question 3. The general characteristics extracted consisted of authors, year of publication, and sample information about participants' number and types (e.g., healthy people or stroke patients).

2.4. Quality assessment

The methodological quality and potential bias of each research article were evaluated using the QUADAS-2 risk of bias tool, which assesses diagnostic accuracy studies across four key domains: patient selection, index test, reference standard, and flow and timing. The tool uses signaling questions to guide the evaluation process, offering flexibility to customize domains based on the specific requirements of the review [21]. Following the recommendations in [21], additional criteria from a previous review [13] were incorporated into the appraisal process, resulting in the inclusion of two extra domains: statistical analysis and outcome measures specification. The statistical analysis domain examined whether validity indices were appropriately calculated and reported, whether the statistical methods matched the data type, and whether results were presented comprehensively with precision estimates. The outcome measures specification domain focused on the clarity of gait parameter definitions, explicit descriptions of calculation methods, and standardization of the measurement process. Two independent reviewers (XC and YJ) conducted the quality assessment of the included studies. Disagreements were resolved through discussion to achieve consensus.

3. Results

3.1. Literature search results

Detailed information about the screening and eligibility assessment is shown in Fig. 1. The initial search identified a total of 644 articles and 485 studies remained after duplicates were removed. After full-text

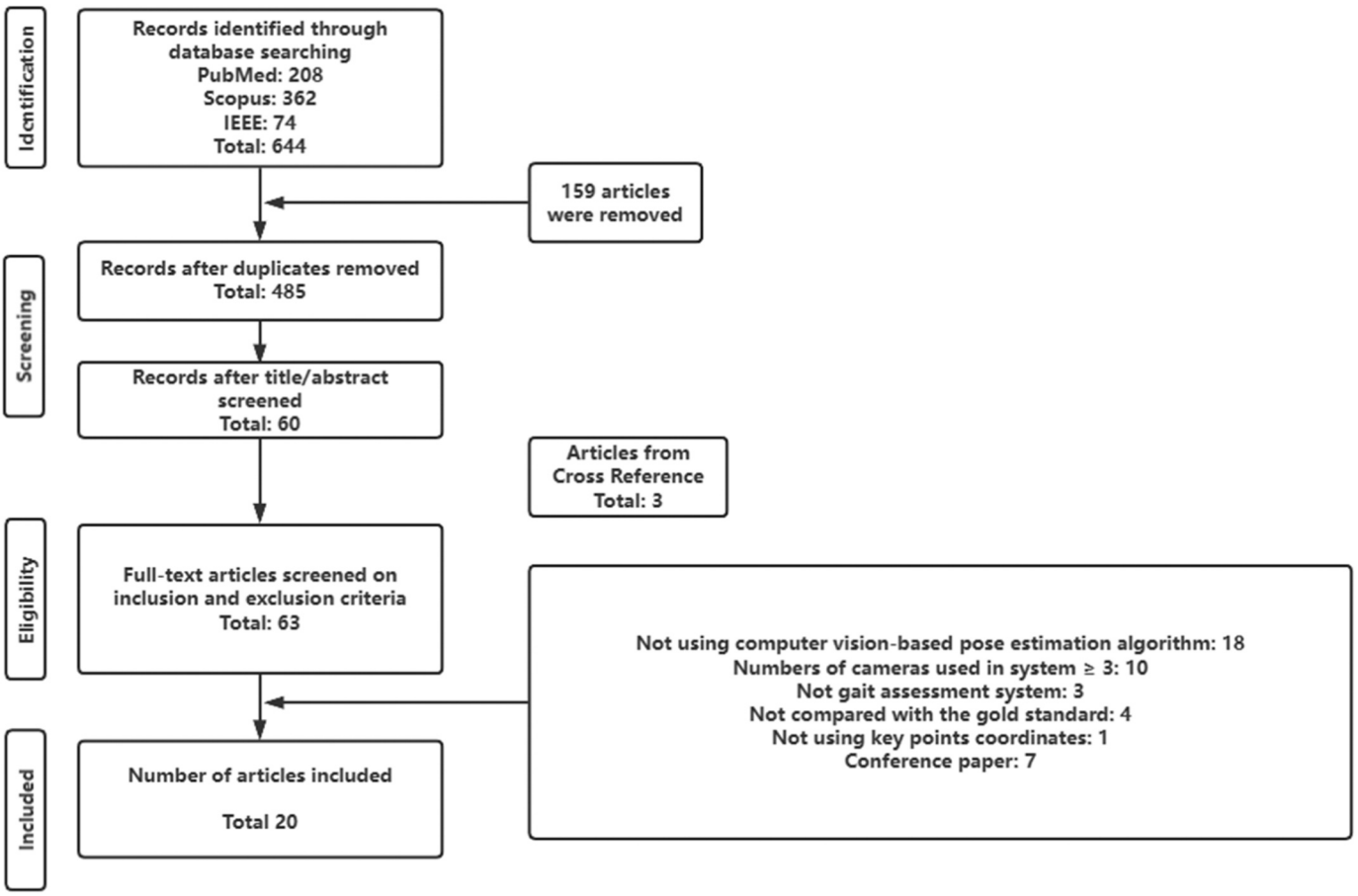

Fig. 1. Flow chart of the selection process of articles in this review.

screening, 20 studies were finally identified that met the inclusion and exclusion criteria.

3.2. Study characteristics

Data on the year of publication, participant characteristics, gait protocol used, and the gold standard comparator, are summarized in Table 2. All included studies were published in the last 5 years, reflecting the rapid development of computer vision technology in gait assessment. The participants involved were not only healthy people but also people with stroke [22], Parkinson's [16], and developmental dysplasia of the hip [24]. The accuracy of the algorithms in overground or treadmill gait assessments was compared to marker-based motion capture systems, including the Opti-Track [1,23], Vicon [5,8,9,16,22,24–32], Smart-D [33], MAC3D [34], and the Motion Analysis Corp [14], except for 2 studies that used the force plate [14] and the pressure sensor [15] as the gold standard comparator. Outcome measures included spatiotemporal parameters such as gait length and gait speed and the joint kinematic parameters involving the hip, knee, and ankle joint angles (as shown in Table 2). Only one research article also examined the accuracy of joint kinetic parameters including the ground reaction force and knee adduction moment by scaling a musculoskeletal model to the participant's anthropometry using OpenSim's Scale tool [14].

3.3. Quality assessment

The quality assessment results are presented in Table 3. Only one study achieved low bias across all domains [5], primarily because most studies were rated as high risk or unclear in the Participant Selection domain. This was largely due to insufficient reporting of sampling methods or recruitment criteria for gait participants in reliability and validity studies, which introduced potential bias. Half of the articles clearly defined outcome measures and calculation methods. Low bias was observed in the Index Test and Reference Standard domains for 19 studies and in the Flow and Timing domains for 13 studies. Additionally, 17 studies were rated as low risk for applying appropriate statistical methods and multiple measures to assess reliability and validity.

3.4. Reliability and validity evaluation

Table 4 shows the results of the reliability and validity evaluation. Among the included articles, 14 out of 20 evaluated the reliability or validity of OpenPose on gait analysis. Following OpenPose, BlazePose was the next most commonly used, followed by OpenCap, MoveNet, DeepLabCut, KAPAO, and ThreeDPoseTracker. As the main purpose of this article is not to discuss the influence of camera sides (i.e., exposed or occluded sides) on the data accuracy, only data for the limb without occlusion is synthesized.

3.4.1. Reliability

Of the reviewed articles, three evaluated the test-retest reliability of OpenPose and KAPAO using the Intraclass Correlation Coefficient (ICC) and Standard Error of Measurement (SEM), focusing on kinematic parameters, including segments and joint angles. The levels used for the interpretation of ICC values in this study follow the threshold proposed by Koo and Li [35], where values below 0.5 indicate poor reliability, 0.5–0.75 indicate moderate reliability, 0.75–0.9 indicate good reliability, and values above 0.9 indicate excellent reliability. Specifically, OpenPose demonstrated good to excellent reliability for hip, knee, and ankle angles in the sagittal plane with ICCs ranging from 0.77 to 0.99.

Table 2
Characteristics of the identified studies.

Study	Algorithm	Participant type (Number of participants)	Gold standard	Walking protocol	Gait parameters
Zago et al.; 2020 [33]	OpenPose	Healthy (2)	8-camera Smart-D, BTS (100 Hz)	Overground	Spatiotemporal
Yamamoto et al.; 2021 [1]	OpenPose	Healthy (11)	7-camera Opti-Track (100 Hz)	Treadmill	Spatiotemporal; Joint kinematics
John et al.; 2024 [22]	OpenPose	Stroke (30)	10-camera Vicon (ND)	Overground; Treadmill	Spatiotemporal; Joint kinematics
Stenum et al.; 2024 [16]	OpenPose	Stroke (44); PD (19)	10-camera Vicon (100 Hz)	Overground	Spatiotemporal; Joint kinematics
Stenum et al.; 2021 [9]	OpenPose	Healthy (31)	10-camera Vicon (100 Hz)	Overground	Spatiotemporal; Joint kinematics
Takeda et al.; 2020 [34]	OpenPose	Healthy (1)	8-camera MAC3D (60 Hz)	Treadmill	Joint kinematics
Hii et al.; 2023 [5]	BlasePose	Healthy (31)	10-camera Vicon (100 Hz)	Overground	Spatiotemporal
Lin et al.; 2023 [24]	KAPAO: YOLO v5	Healthy (18); Developmental dysplasia of the hip (10)	8-camera Vicon (100 Hz)	Overground	Joint kinematics
Washabaugh et al.; 2022 [25]	MoveNet Lightning/ MoveNet Thunder/ DeepLabCut	Healthy (31)	10-camera Vicon (100 Hz)	Overground	Spatiotemporal; Joint kinematics
Talaa et al.; 2023 [26]	OpenPose	Healthy (32)	10-camera Vicon (100 Hz)	Overground	Joint kinematics
Horsak et al.; 2023 [27]	OpenCap	Healthy (19)	16-camera Vicon (120 Hz)	Overground	Joint kinematics
Aoyagi et al.; 2022 [8]	ThreeDPoseTracker	ND	ND-Vicon (ND)	Overground	3D coordinates of joints
Menychtas et al.; 2023 [28]	BlasePose	Healthy (17)	10-camera Vicon (100 Hz)	Treadmill	Joint kinematics
Yang et al.; 2024 [23]	BlasePose	Healthy (20)	9-camera Optitrack (120 Hz)	Treadmill	Joint kinematics
Boldo et al.; 2024 [29]	OpenPose	Healthy (8)	8-camera Vicon (120 Hz)	Overground	Joint kinematics
Uhlrich et al.; 2023 [14]	OpenCap	Healthy (10)	8-camera Motion Analysis Corp (100 Hz) and in-ground force plates (Bertec Corp, 2000 Hz)	Overground	Joint kinematics and kinetics
Shin et al.; 2021 [15]	OpenPose	PD (16); Healthy (15)	GAITRite (pressure sensor)	Overground	Spatiotemporal
Liang et al.; 2022 [30]	OpenPose	Healthy (30)	6-camera Vicon (60 Hz)	Treadmill	Joint kinematics
Ino et al.; 2023 [31]	OpenPose	Healthy (21)	6-camera Vicon (120 Hz) and OR6 force plates (1200 Hz)	Overground	Joint kinematics
Ota et al.; 2021 [32]	OpenPose	Healthy (24)	8-camera Vicon (60 Hz)	Treadmill	Joint kinematics

ND: not mentioned; N/A: not applicable; PD: Parkinson's disease; CP: cerebral palsy; UPDRS III: Unified Parkinson's Disease Rating Scale part III

Table 3
Results of risk of bias assessment.

Study	Participant Selection	Outcome Measures Specification	Index Test	Reference Standard	Flow and Timing	Statistical Analysis
Zago et al.; 2020 [33]	1	0	0	0	0	0
Yamamoto et al.; 2021 [1]	2	0	0	0	0	0
John et al.; 2024 [22]	1	1	0	0	1	0
Stenum et al.; 2024 [16]	1	0	0	0	0	0
Stenum et al.; 2021 [9]	1	0	0	0	0	0
Takeda et al.; 2020 [34]	1	0	0	0	0	0
Hii et al.; 2023 [5]	0	0	0	0	0	0
Lin et al.; 2023 [24]	1	2	0	0	0	0
Washabaugh et al.; 2022 [25]	1	0	0	0	0	0
Talaa et al.; 2023 [26]	1	1	0	0	2	1
Horsak et al.; 2023 [27]	2	2	0	0	0	0
Aoyagi et al.; 2022 [8]	1	1	2	0	1	1
Menychtas et al.; 2023 [28]	1	1	0	0	2	0
Yang et al.; 2024 [23]	2	0	0	0	0	0
Boldo et al.; 2024 [29]	1	0	0	0	0	0
Uhlrich et al.; 2023 [14]	1	2	0	0	2	0
Shin et al.; 2021 [15]	0	2	0	1	2	0
Liang et al.; 2022 [30]	1	1	0	0	2	0
Ino et al.; 2023 [31]	2	1	0	0	0	0
Ota et al.; 2021 [32]	2	0	0	0	0	1

0 = low risk; 1 = high risk; 2 = unclear risk

Table 4
Characteristics of pose algorithms and their reliability and validity.

Study	Type of algorithm	Camera settings and data dimension	Reliability Performance	Validity Performance
Yamamoto et al. [1]	OpenPose-BODY_25 pose	Monocular; 2D	N/A	Outcome measures: MAE (mean ± sd) [°] / ICCs / CCC Spatiotemporal (gait speed/ stride/ stride time/ step length): 0.01 ~ 0.04 / 0.88 ~ 0.98 / NA; Hip F/E angle: 3.91 ~ 4.47 / 0.87 ~ 0.91 / 0.99; Knee F/E angle: 2.17 ~ 3.66 / 0.41 ~ 0.93 / 0.99; Ankle F/E angle: 6.87 ~ 7.62 / 0.01 ~ 0.60 / 0.93
John et al. [22]	OpenPose-BODY_25 pose	Monocular; 2D	N/A	Outcome measures: ICC1,1 / CCC Spatiotemporal: 0.745 ~ 0.994 / NA; Hip F/E: 0.682 / 0.96; Knee F/E: 0.94 / 0.99; Ankle F/E: 0.173 / 0.77
Stenum et al. [16]	OpenPose-BODY_25 pose	Monocular; 2D	N/A	Outcome measures: MAE [°] / r Spatiotemporal: Step time [s]: 0.028 / 0.997; Step length [m]: 0.028 / 0.967; Gait speed [m/s]: 0.04 / 0.981; Step time asymmetry: 0.03 / 0.977; Step length asymmetry: 0.05 / 0.890; Kinematic: Trunk inclination angle [°]: 1.5 (−3.0–2.9) / 0.964; Hip F/E: 2.7 ~ 3.3; Knee F/E [°]: 3.5 ~ 4.0; Ankle F/E: 4.8 ~ 6.3
Stenum et al. [9]	OpenPose-BODY_25 pose	Monocular; 2D	N/A	Outcome measures: MAE [°]/ r / ICC Temporal (s): 0.00 ~ 0.011 / 0.909 ~ 0.998 / 0.963 ~ 0.998; Spatial: 0.02 ± 0.02 / 0.973 / 0.970; Spatiotemporal (m/s): 0.02 ± 0.02 / 0.986 / 0.985; Joint kinematic measures: MAE [°] / CCC Hip F/E: 3.7 ± 2.0 / 0.979 ± 0.022; Knee F/E: 5.1 ± 2.1 / 0.984 ± 0.012; Ankle F/E: 6.3 ± 3.4 / 0.778 ± 0.303
Boldo et al. [29]	OpenPose-BODY_25 pose	Monocular; 2D	N/A	Joint kinematic measures: R2 / Bland–Altman mean differences [°] Hip F/E: 0.96 ± 0.03 / (0.2°−1.8°); Knee F/E - ES: 0.95 ± 0.03 / (1.2°−6.1°); Ankle PF/DF: 0.85 ± 0.08 / (1.2°−7.4°)
Talaa et al. [26]	OpenPose-BODY_25 pose	Monocular; 2D	N/A	Outcome measures: MAE [°] Right knee F/E: 4.9°± 0.96; Left knee F/E: 5.2°± 1.83
Ino [31]	OpenPose-BODY_25 pose	Monocular; 2D	Test-retest reliability: ICC 1,3 Ankle PF/DF: 0.77–0.93; Knee F/E: 0.81–0.99; Hip F/E: 0.92–0.9	Joint kinematic measures: MAE [°] / Pearson's r / CMC Ankle PF/DF: 3.1 (2.7–3.5) / 0.734 ~ 0.877 / 0.936; Knee F/E: 2.3 (2.1–2.6) / 0.806 ~ 0.987 / 0.994; Hip F/E: 2.5 (1.1 – 3.9) / 0.913 ~ 0.932 / 0.945;
Menychtas et al. [28]	OpenPose-BODY_25 pose	Monocular; 2D	N/A	Joint kinematics measures: Bland-Altman mean difference (95 %LOA) [°] Hip F/E: −5.84(−16 ~ 4.6); Knee F/E: −11.14 (−16 ~ −6.2); Ankle PF/DF: −9.77 (−20 ~ 0.72);
Washabaugh et al. [25]	OpenPose-BODY_25 pose	Monocular; 2D	N/A	Joint kinematic measures: MAE [°] Knee F/E: OpenPose: 5.1 ± 2.5; Hip F/E: OpenPose: 3.7 ± 1.3
Ota [32]	OpenPose-BODY_25 pose	Monocular; 2D	N/A	Joint kinematic measures: R^2 / $ICC_{2,1}$ / Bland-Altman [°] Frontal plane: Pelvis elevation/depression: 0.002 ~ 0.139 / −0.014 ~ 0.069 / −4.111 ~ 3.807; Hip abduction / adduction: 0.080 ~ 0.533 / 0.042–0.473 / −5.867 ~ 0.584; Sagittal plane: Hip F/E: 0.158 ~ 0.678 / 0.656 ~ 0.864 / −3.092 ~ 1.449; Knee F/E: 0.707 ~ 0.883 / 0.670 ~ 0.914 / 3.140 ~ 6.869; Ankle PF/DF: 0.433 ~ 0.663 / 0.465 ~ 0.620 / −6.604 ~ 3.123
Takeda et al. [34]	OpenPose-BODY_25 pose	Monocular; 2D	N/A	Joint kinematic measures: $ICC_{2,1}$ Hip F/E: 0.97; Knee F/E: 0.92; Ankle PF/DF: 0.57
Washabaugh et al. [25]	MoveNet Lightning, MoveNet Thunder, and DeepLabCut	Monocular; 2D	N/A	Joint kinematic measures: MAE [°] Knee F/E: OpenPose : 5.1 ± 2.5; MoveNet Thunder: 7.5 ± 2.5; MoveNet Lightning: 9.1 ± 3.0; Deep Lab Cut: 9.4 ± 2.4; Hip F/E: OpenPose: 3.7 ± 1.3; MoveNet Thunder:.6 ± 1.; MoveNet Lightning: 5.9 ± 3.6; DeepLabCut: 6.8 ± 1.6
Lin et al. [24]	KAPAO: adapted from YOLO v5	Monocular; 2D	Test-retest reliability Outcome measures: ICC / SEM; Hip F/E: 0.736–0.793 / 2.03–2.91; Knee F/E: 0.752–0.894 / 1.26–2.07	Joint kinematic measures: ICC / SEM / Bland-Altman [°] Hip F/E: 0.929 ~ 0.957 / 0.87 ~ 1.48 / −1.04 ~ 0.51; Knee F/E: 0.835 ~ 0.921 / 1.05 ~ 1.71 / −1.42 ~ 0.46
Tony Hii et al. [5]	BlasePose GHUM Heavy (MediaPipe Pose)	Monocular; 3D (only 2D were used)	N/A	Spatiotemporal measures: MAE [°] / ICC2,1 / r Temporal: 0.03 ~ 0.05 / 0.469 ~ 0.867 / 0.522 ~ 0.9
Menychtas et al. [28]	BlasePose GHUM Heavy (MediaPipe Pose)	Monocular; 3D (only 2D were used)	N/A	Joint kinematics measures: Bland-Altman mean difference (95 %LOA) [°] Hip F/E: −26.25 (−56 ~ 3.9); Knee F/E: 17.75(−35 ~ −0.55); Ankle PF/DF: 23.8 (5.5 ~ 42)
Yang et al. [23]	BlasePose GHUM Heavy (MediaPipe Pose)	Monocular; 3D (only 2D were used)	N/A	Joint kinematic measures: MAE [°] / r Camera view angle: 90°: Hip F/E: 1.95 ± 0.34 / 0.96

(*continued on next page*)

Table 4 (*continued*)

Study	Type of algorithm	Camera settings and data dimension	Reliability Performance	Validity Performance
				± 0.026; Knee F/E: 2.58 ± 0.79 / 0.98 ± 0.018; Ankle PF/DF: 2.19 ± 0.4 / 0.84 ± 0.74
Shin et al. [15]	OpenPose-BODY_25 pose	Monocular; 3D	N/A	Spatiotemporal: ICC (95 % interval) 0.90 ~ 0.98
Liang et al. [30]	OpenPose-BODY_25 pose	Monocular; 3D	Test-retest reliability: ICC (C, 1) Upper Trunk 3D Segmental Angle: 0.603–0.714; Thigh 3D Segmental Angle: 0.506–0.617; Shank 3D Segmental Angle: 0.519–0.652 Knee F/E angles: 0.734;	Joint kinematic measures: $ICC_{A,1}$/ Bland-Altman mean (approximate value from graph) [°] Knee F/E: 0.726 / −5°; Upper Trunk F/E: 0.694 / −6°; Thigh F/E: 0.716 / −9°; Shank F/E:0.760 / −0.5°;
Aoyagi et al. [8]	ThreeDPoseTracker	Monocular; 3D	N/A	Pearson's correlation coefficient for each coordinate (X, Y, Z) Right Hip joint(0.85, −0.02, −0.85); Left Hip joint (0.79, 0.52, −0.29); Right Knee(0.93, −0.77, −0.60); Left Knee (0.77, 0.49, −0.19); Right Ankle (0.84, −0.75, −0.88); Left Ankle (0.81, 0.14, −0.61); Right Toe (0.92, −0.66, −0.77); Left Toe (0.74, 0.57, −0.72)
Zago et al. [33]	OpenPose-COCO	Binocular; 3D	N/A	RMSE: step length (1.53 ~ 1.93 cm); swing time (0.02 ~ 0.05 s); stance time (0.05 s); skeleton nodes (20.8 ± 3.5 mm)
Uhlrich et al. [14]	OpenCap	Binocular; 3D	N/A	Outcome measures: RMSE [°] Anatomical marker error[mm]: 31 ~ 35; Kinematics-Rotations(pelvis tilt / list /rotation / hip flexion / abduction /rotation / knee flexion / ankle dorsiflexion / subtalar angle / lumbar extension / bending / rotation)[°]: 4.1 (2.3 ~ 6.6); Kinematics-Translations[mm]: 12.3 (6.8 ~ 19.6); Ground reaction forces [%BW]: 1.1 ~ 8.2 (6.7 % ~ 17.1 %); Joint moments (hip flexion/adduction/rotation moment; knee extension moment; ankle plantarflexion moment; subtalar angle moment; lumbar extension/bending/ rotation moment)[%BW*height]: 0.75 (0.20–1.32,19 %)
Horsak et al. [27]	OpenCap	Binocular; 3D	N/A	Outcome measures: RMS [°] Pelvis (tilt / list / rotation): 3.0 ~ 6.0; Hip (flexion / adduction / rotation): 5.4 ~ 7.6 / 3.7 ~ 4.8 / 6.6 ~ 8.4; Knee F/E: 5.7 ~ 8.5; Ankle PF/DF: 4.6 ~ 7.9; Subtalar angle: 7.3 ~ 10.2

RMSE: Root Mean Square Errors; LOA: Limit of Agreement; TDPT: ThreeDPoseTracker; LFM: Linear Fit Method; MAE: mean absolute error; AE: absolute error; CI: confidential interval; ICC: intraclass correlation coefficient; ROM: range of motion; CCC: cross-correlation coefficients; Sag: sagittal plane; Fron: frontal plane; SEM: Standard error of measurement; F/E: Flexion/Extension; PF/DF: Plantarflexion/Dorsiflexion.

For segment angles, the knee showed moderate reliability with ICC = 0.734, followed by trunk, shank, and thigh angles (ICCs = 0.506–0.714) [30]. KAPAO also showed moderate to good reliability for lower extremity joint kinematics (ICCs = 0.736–0.894) with SEMs under 3° and no significant differences observed [24].

3.4.2. *Validity*

In the reviewed articles, spatiotemporal, kinematic, and kinetic parameters were assessed using various statistical tests, including ICCs, Concordance Correlation Coefficient (CCC), Bland-Altman with Limits of Agreement (LOA), Pearson correlation coefficient, and Mean Absolute Error (MAE). Seven studies [1,5,9,15,16,22,33] focused on spatiotemporal parameters including step length, step length asymmetry, swing time, stance time, gait speed, stride length, stride time, step time, step time asymmetry, and double support time. Kinematic parameters included knee, hip, and ankle angles in the sagittal plane [1,9,16,22–26,28,29,31,32,34], along with trunk inclination, hip abduction/adduction/ rotation, pelvis tilt/rotation, and lumbar angles [9,14,27,32]. Some studies also validated kinetic parameters, including ground reaction forces and joint moments, and compared skeleton node accuracy to the gold standard [8,14,33].

The validity of spatiotemporal parameters was assessed for OpenPose [1,9,15,16,22,33] and BlazePose [5]. Step length, step time, step time asymmetry, gait speed, stride length, and stride time showed good to excellent validity (ICCs = 0.89–0.994; r = 0.886–0.997). However, BlazePose showed poor validity for swing time (ICCs = 0.510) and double support time (ICCs = 0.469) compared to OpenPose (ICCs = 0.909; 0.963).

Current evidence shows that binocular algorithms demonstrated relatively high validity for 3D joint kinematics, with OpenCap exhibiting lower errors in joint rotational angles and translations, but further validation is needed for clinical applications. The MAEs of OpenCap for 18 rotational degrees of freedom (ankles [2 × 2], knees [2 × 1], hips [2 ×3], pelvis rotations [3], lumbar [3] and three translational degrees of freedom (pelvis translations in x, y, and z-axis) were 4.1° (range = 2.3–6.6°) and 12.3 mm (6.8–19.6 mm). Zago et al. [33] reported a lower skeleton node error of 20.8 ± 3.5 mm with a two-camera setup based on OpenPose compared to OpenCap's anatomical marker error of 31–35 mm. Monocular algorithms utilizing deep learning algorithms to estimate 3D position data exhibited lower validity, with overall ICCs ranging from 0.255 to 0.665. For 2D data captured by monocular algorithms, accuracy varied across different joints and algorithms. For OpenPose, hip and knee joint angles in the sagittal plane throughout the gait cycle showed moderate to excellent agreement (ICC = 0.656–0.92) with minimal angle differences (MAE: 2.7–5.2°) and strong waveform agreement (CCC = 0.9–0.994) with gold standard. While ankle kinematics had lower agreement and consistency (ICC: 0.173–0.57; MAE: 3.1–9.77°; CCC: 0.51–0.936). KAPAO showed good to excellent agreement (ICCs = 0.853–0.957; Bland-Altman = −1.42–0.51°), but its validity has only been assessed in a single study. MoveNet and DeepLabCut had MAEs ranging from 5.9° to 7.5°. The validity of ThreeDPoseTracker remains unclear as it only assessed the correlation of joint positions with

Vicon data. BlazePose showed significant inconsistency in its accuracy, with one study reporting joint errors around 30° [28], while another found MAE of less than 3° [23].

OpenCap is the only system capable of estimating kinetic parameters, including ground reaction forces in the vertical, anterior-posterior, and mediolateral directions. Results showed averages of 8.2 % body weight (BW), 2.1 % BW, and 1.1 % BW in vertical, anterior-posterior, and mediolateral ground reaction forces. Joint moments (e.g., knee flexion and ankle plantarflexion moments) are reported as 0.75 % BW*height, ranging from 0.20 to 1.32 (Table 4)[14].

3.5. Specifications of PEAs used for gait analysis

Eight PEAs and their variants were identified in the research (Table 5), with keypoints captured varying across algorithms. OpenCap captured the most keypoints (43), while DeepLabCut captured the fewest (14), excluding feet. Most algorithms provided 2D joint position data via a single camera (i.e., direct output from the algorithm, while 3D data was obtained in some studies through depth estimation (i.e., joint coordinates in the z-axis) using the PNP algorithm from the open-source OpenCV platform [15], 3DPoseNet algorithm [30], and Heatmap [8]. Binocular setups were also used for 3D reconstruction through triangulation of synchronized 2D video keypoint positions [14,33]. The adopted backbone architectures of CNNs varied among the reviewed algorithms and influenced inference times, with BlazePose and MoveNet utilizing lightweight MobileNet V2 backbone for faster CPU inference times (33.3 ms [36] and 100 ms [37]) (Table 5), compared to deeper backbones like ResNet [8,24], VGG-19 [1,22], and HRNet [14] (used by OpenPose and ThreeDPoseTracker), optimized for GPU use. Public datasets, including COCO, MPII, and CrowdPose, were the most commonly used across the reviewed algorithms, while customized datasets, such as those for foot [6], Blaze face, Blaze palm [37], and Active [36], were also used to improve algorithm performance.

4. Discussion

This systematic review aimed to evaluate the reliability and validity of current PEA-based gait analysis and synthesize algorithmic factors affecting accuracy. The accuracy of the algorithms varied, with some demonstrating precision that approaches the standards required for clinical gait analysis in 2-D sagittal plane analysis, primarily in typically developing populations. However, further validation is necessary to confirm their applicability across other planes and pathological populations. Key factors associated with accuracy included backbone architecture, training datasets, and the number of cameras used. The following sections will discuss these components.

4.1. Reliability and validity of PEA-based gait analysis

4.1.1. Reliability

OpenPose and KAPAO demonstrated moderate to excellent reliability on joint angles (ICCs = 0.5–0.99). However, the limited body of evidence and lack of comparison with other algorithms make it challenging to draw definitive conclusions. Notably, while the ICC (0.77–0.99) of OpenPose falls into good to excellent according to the commonly used threshold [35], an ICC of 0.77 may still be insufficient for clinical applications and may result in considerable between-session variability at the individual level. Bland-Altman plots may provide valuable insights into measurement variability in clinical applications and could be useful for further validation of OpenPose's reliability [39]. Medical practitioners need to carefully consider changes in joint angles when assessing disease progression.

4.1.2. Validity

In collated studies, validity varied across algorithms, with binocular PEA showing higher accuracy than monocular algorithms in spatio-temporal [33] and joint kinematic parameters [14]. Among binocular PEA systems, OpenCap demonstrated a mean Root Mean Square Error (RMSE) of 4.8° (range: 2.6°–7.4°) in lower-extremity gait analysis compared to a marker-based motion capture system in healthy individuals [14]. This range overlaps with reported RMSE values from the 8-camera-based Theia 3D system (2.6°–11° for walking, running, and cycling) [40,41] and inertial measurement units (2.0°–12°) for walking, running [42,43], suggesting a similar order of magnitude in error. However, it should be noted that direct statistical comparisons are currently unavailable, and further studies are needed to confirm its potential superiority. Clinically, joint kinematic errors of 2° or less are widely acceptable, while errors between 2° and 5° are reasonable but may need careful interpretation [13]. Horsak et al. found a slightly higher overall RMSE of 5.8 ± 1.8° for OpenCap, deeming it too large for clinical use [27]. this 1-degree difference may be due to errors in the marker-based motion capture system, such as marker placement variation, kinematic crosstalk, and soft tissue artifacts [40], with reported soft tissue errors averaging 4–8° in hip joint motion during activities like walking [44], and up to 3° in knee joint angles [45]. Careful validation and comparison with gold standards like bi-planar video radiography, as

Table 5
Algorithm specifications of PEAs.

Type of algorithms	No. of camera	Backbone architecture	No. of Keypoints	Dimension of position data	Database	Inference Time:
OpenPose-COCO	2	VGG−19	18	3D: Triangulation of 2D OpenPose data [33]	COCO	High in GPU; Low in CPU [6]
OpenPose-Body_25	1/2	VGG−19	25	2D: Output from OpenPose; 3D: depth data estimated using a PNP algorithm from an open-source (OpenCV) platform [15] / estimated using 3DPoseNet [30]	COCO; MPII; customized foot dataset [6]	High in GPU; Low in CPU [6]
BlasePose GHUM (MediaPipe Pose)	1	MobileNetV2	33	3D: depth data estimated from the GHUM model using 2D projections [5,23]	COCO; Blaze face; Blaze palm [37]	Low in CPU [37]
KAPAO (adapted from YOLO v5)	1	CSPDarknet53	18	2D	COCO; CrowdPose [38]	-
DeepLabCut	1	ResNet 50	14	2D	-	-
ThreeDPoseTracker	1	ResNet34	24	3D: depth data estimated using Heatmap	COCO; customized 3D pose tracking dataset	Low in GPU [8]
OpenCap: OpenPose and HRNet [14]	2	VGG−19/HRNet	43	3D: Triangulation of the synchronized 2D video keypoint positions [14]	COCO; MPII; customized foot dataset [6]	Low in GPU
MoveNet	1	MobileNetV2	17	2D	COCO; Active (internal Google dataset) [36]	Low in CPU [36]

Keypoints: VGG-19: Visual Geometry Group 19 Layer CNN; ResNet 50: Residual Network 50 layer CNN; HRNet: High-Resolution Net; COCO: Common Objects in Context dataset; PNP: perspective-n-point; 3D PoseNet: a 3D keypoint detection deep learning algorithm

suggested by Wade et al., are necessary to ensure the validity of OpenCap for high-precision applications [46].

Among the monocular PEA systems reviewed, the error varied depending on the movement plane and the specific joints involved. For example, the hip joint angles from OpenPose in the sagittal plane throughout the entire gait cycle showed moderate to excellent agreement (ICC = 0.656–0.9), with minimal joint angle differences (MAE: 2.7–3.7°) and excellent waveform agreement with the gold standard (CCC = 0.9–0.979). Knee joint angles follow, with ICC values ranging from 0.67 to 0.92, MAE between 3.5 and 5.2°, and CCC from 0.978 to 0.994. However, performance in ankle kinematics is poor, with large deviations (ICC: 0.173–0.57; MAE: 3.1–9.77°; CCC: 0.51–0.936). This issue also exists in other algorithms, such as BlazePose, because the foot is a highly complex segment, presenting ongoing challenges even for marker-based systems [27]. Clinical practitioners should be cautious when interpreting changes in the ankle joint. Other algorithms have shown even lower accuracy for joint angles in the sagittal plane compared to OpenPose. Therefore, if accuracy is a key requirement for developing a gait analysis system based on monocular PEA, OpenPose would be the recommended choice [25].

4.1.3. Clinical utility

Among the reviewed PEA algorithms, binocular solutions (e.g., OpenCap) show promise for certain clinical applications, given their relatively high accuracy in the sagittal plane. For example, OpenCap utilized a Long Short Term Memory neural network model to predict the 3D positions of 43 anatomical keypoints based on the 3D positions of 20 video keypoints and achieved anatomical keypoint errors within 35 mm [14]. This means kinematics of all degrees of freedom of the body segments can be defined, allowing for more clinically meaningful joint kinematic measures. Additionally, OpenCap is able to obtain the joint kinetic data with reasonable accuracy (as shown in Table 4.) using OpenSim, such as knee adduction moment [14], offering clinicians convenient access to critical kinetic parameters.

Monocular PEAs, while not designed to match the accuracy of marker-based systems, offer more accessible solutions for scenarios where simplicity and scalability are prioritized in clinical practice [16]. Using a normal RGB camera, these algorithms identify and track keypoints, simplifying setup and allowing large-scale data collection for gait feature extraction and data mining [47]. Although monocular PEAs are primarily designed for convenience rather than precision, they demonstrate a balance between accuracy and ease of use for abnormal gait detection, disease severity classification, and early disease detection [17,18,48–51] by integrating with classification models such as the Random Forest classifier [19].

4.2. Algorithmic factors associated with accuracy of PEA

The accuracy of pose estimation algorithms is influenced by various factors, such as video resolution, camera orientation, occlusion, and lighting. Since these external factors have been reviewed elsewhere [46,52], our focus is on algorithmic factors affecting accuracy.

4.2.1. Camera settings and data dimension

Camera settings, particularly the number of cameras, can influence motion capture accuracy. For vision-based systems, occlusion is a major detrimental factor [33]; when keypoints are obscured, the algorithm will lose them or estimate their actual positions, significantly deteriorating the data quality [33]. Multiple-camera setups reduce errors caused by occlusion by capturing multi-view information [33,46]. Besides, camera numbers also determine the dimension of joint position data. Monocular PEAs typically produce 2D-pixel coordinates [6] or the relative coordinates to the center of the body [8], derived from keypoint heatmaps that represent the likelihood of each pixel being a joint [6,53]. While most reviewed studies use 2D data for gait analysis, 3D data can be estimated through post-processing on 2D outputs [54], as applied in studies [5,8,15,30]. For a detailed review of methods to estimate 3D position data from monocular pose estimation algorithms, see the previous study [54]. However, this method usually suffers from inaccuracy due to a lack of direct depth information [55], which explains the low validity of depth data in the reviewed studies. Alternatively, 3D coordinates can be obtained by triangulating 2D coordinates from multiple views, as done by OpenCap, providing key advantages over monocular vision, particularly in improving 3D data accuracy and handling occlusions.

4.2.2. Backbone architecture

In deep learning and computer vision, a backbone is a pre-trained CNN that serves as the foundational structure for processing input images and extracting features used by subsequent layers for tasks like object detection, segmentation, or pose estimation [56]. Research shows that deeper backbone networks can extract more discriminative features, improving the task accuracy [6,57]. VGG, introduced in 2014 with only 19 layers, was an early backbone used by OpenPose [58]. By contrast, ResNet, suggested in 2016, increased the depth up to 152 layers by adding the residual networks and improved accuracy by 28 % on the COCO dataset [57]. HRNet, used in OpenCap, is also a complex backbone to increase accuracy by maintaining high-resolution features throughout the entire network, enabling it to capture fine-grained image details [59]. However, a complex backbone results in increased computational complexity and memory usage, limiting its suitability for certain applications. CSPDarkNet, introduced in 2020 for YOLO, achieved an effective balance between accuracy and computational efficiency by reducing computation by 20 %, making it ideal for real-time applications [60]. MobileNet, adopted in MoveNet and MediaPipe BlazePose, is an extremely lightweight backbone, optimized for mobile devices, offering reduced latency and faster inference speeds, though at the cost of lower accuracy [61]. To achieve higher accuracy, computational efficiency often has to be compromised. Therefore, when accuracy is prioritized, models with more complex backbones could be emphasized.

4.2.3. Training datasets

In pose estimation algorithms, datasets are crucial to accuracy, as they provide annotated images for training models to identify key joints. PEAs typically use CNNs to process images and extract high-level features. During training, the model learns to identify keypoints (e.g., shoulders, elbows, knees) from annotated images and predicts joint locations by analysing patterns of pixel color, gradient, and texture. The training process optimizes model parameters to minimize errors between predicted joints and annotated ground truth. The accuracy of the model largely depends on the quality and diversity of the dataset, which determines how well the model generalizes and the number of key points it can predict [46]. Utilizing high-quality, diverse, and accurately annotated datasets is key to improving the accuracy of PEAs [54].

The most commonly used datasets to train 2D pose estimation algorithms, such as COCO and MPII, are large-scale public datasets [54]. However, these datasets predominantly feature normal daily activities and static poses and lack sufficient samples of pathological gait patterns [54]. Even for datasets that include gait patterns, such as the foot dataset used by OpenPose [7] or the Joint-annotated Human Motion Database featuring sports movement, the diversity is insufficient, failing to cover the wide range of abnormal gait variations. This limitation results in poor generalization of the model for gait pattern recognition. Additionally, many datasets provide limited keypoint annotations, excluding critical points like foot keypoints, necessary for clinical diagnosis of gait abnormalities (e.g., hemiplegic, spastic, and Parkinsonian gait). For instance, COCO lacks foot keypoints, making it unsuitable for analysing ankle kinematics in models like YOLO or OpenPose-COCO [62]. Finally, the quality of keypoints annotations also significantly affects the model accuracy, especially when occluded joints are inaccurately labeled, which may mislead the model into learning biomechanically incorrect

joint locations. Inconsistent or inaccurate annotations, often due to a lack of professional biomechanical expertise, further degrade model accuracy [46].

This issue can be addressed by using specific datasets from the start of model training or employing transfer learning to fine-tune pre-trained models on new, professionally annotated datasets. For example, OpenPose was trained with a specific foot dataset combined with COCO to identify foot keypoints, which has made it widely used in gait analysis systems [6]. MediaPipe BlazePose also adopted this approach, identifying heel and toe keypoints. In addition, Takumi et al. [63] demonstrated that transfer learning, leveraging a new sports dataset, can significantly improve 2D pose estimators for extreme poses while maintaining high performance on standard datasets. However, creating pathological gait datasets poses significant challenges due to the need for precise manual annotation and the labor-intensive nature of this process. Privacy and ethical concerns also limit the collection and sharing of clinical gait data, making it challenging to obtain the high-quality, annotated datasets needed to improve gait recognition algorithms [46].

4.3. Limitations

This review excluded non-English journals that may have restricted the comprehensive gathering of literature on automatic assessment systems. Besides, this study did not distinguish between healthy individuals and patients when assessing validity and reliability; instead, articles that examined the accuracy of gait analysis for healthy individuals and/or patients were all reviewed. Hence, for certain patient groups, such as those with cerebral palsy, the accuracy of keypoint detection may be lower than the precision summarized in this study when tracking unusual movements, due to factors like occlusions, rapid motion, atypical body poses, or deviations from the data used to train the model. Therefore, future research could focus on reviewing the accuracy of pose estimation algorithms specifically in patient populations.

4.4. Future research

For future studies that use pose estimation algorithms in clinical settings, ensuring the accuracy of these algorithms remains crucial. Researchers could consider employing transfer learning by utilizing hyperparameters from existing well-trained models and developing customized datasets tailored to populations with specific diseases. For clinical applications, integrating a lightweight pose estimation algorithm into web-based platforms could facilitate large-scale data collection. This would allow for comprehensive data mining to uncover detailed patterns and insights, thereby enabling a thorough analysis of patient rehabilitation progress. Such an approach could reveal new correlations and trends that inform and optimize treatment plans. Additionally, there is an increasing interest in integrating multimodal fusion strategies with pose estimation algorithms, which involves incorporating data from various sensors and sources, such as imaging, biosignals, and clinical assessment [64]. This method enhances the accuracy and comprehensiveness of pose estimation by combining information from multiple modalities. By leveraging these diverse inputs, the system can provide a more holistic analysis of patient movement and health, ultimately improving the precision of diagnoses and the effectiveness of rehabilitation programs.

5. Conclusion

In summary, innovative pose estimation algorithms such as OpenPose, BlazePose, and OpenCap demonstrate varying degrees of accuracy, with current evidence suggesting that binocular solutions (e.g., OpenCap) may have higher validity in kinematic and kinetic estimates compared to monocular algorithms. Despite some limitations, including the need for specific datasets and challenges in capturing certain joint angles, these algorithms offer promising supplements to traditional marker-based systems, especially in clinical settings where minimal setup and ease of use are critical. The solutions show great potential to augment clinical diagnosis and monitor diseases such as Parkinson's disease and cerebral palsy through gait analysis, but further robust evidence is needed before they can be widely adopted in clinical practice. Future researchers aiming to improve the accuracy of these algorithms could focus on customized datasets or compensation algorithms. Those interested in clinical applications could integrate algorithms into web-based platforms for large-scale data collection and analysis or combine multimodal fusion strategies with pose estimation algorithms to aid in clinical diagnosis and rehabilitation. This approach will enable a more comprehensive understanding of patient rehabilitation progress and facilitate the development of more effective treatment plans.

CRediT authorship contribution statement

Cheng Xingye: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Jiao Yiran:** Writing – review & editing, Methodology, Investigation, Formal analysis. **Meiring Rebecca M.:** Writing – review & editing, Supervision. **Sheng Bo:** Writing – review & editing. **Zhang Yanxin:** Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis.

Declaration of Generative AI and AI-assisted technologies in the writing process

Statement: During the preparation of this work the author(s) used ChatGPT in order to improve language. After using this tool, the author (s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of Competing Interest

The authors do not have any conflict of interest which could have influenced the results of this work.

Acknowledgments

The authors would like to acknowledge the Chinese Scholarship Council for living expenses support.

References

[1] M. Yamamoto, K. Shimatani, M. Hasegawa, Y. Kurita, Y. Ishige, H. Takemura, Accuracy of temporo-spatial and lower limb joint kinematics parameters using openpose for various gait patterns with orthosis, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 2666–2675, https://doi.org/10.1109/TNSRE.2021.3135879.
[2] A. Cappozzo, Gait analysis methodology, Hum. Mov. Sci. 3 (1984) 27–50, https://doi.org/10.1016/0167-9457(84)90004-6.
[3] A.A. Hulleck, D. Menoth Mohan, N. Abdallah, M. El Rich, K. Khalaf, Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies, Front. Med. Technol. 4 (2022) 901331, https://doi.org/10.3389/fmedt.2022.901331.
[4] T.A.L. Wren, C.A. Tucker, S.A. Rethlefsen, G.E. Gorton, S. Õunpuu, Clinical efficacy of instrumented gait analysis: systematic review 2020 update, Gait Posture 80 (2020) 274–279, https://doi.org/10.1016/j.gaitpost.2020.05.031.
[5] C.S.T. Hii, K.B. Gan, N. Zainal, N. Mohamed Ibrahim, S. Azmin, S.H. Mat Desa, B. van de Warrenburg, H.W. You, Automated gait analysis based on a marker-free pose estimation model, Sensors 23 (2023), https://doi.org/10.3390/s23146489.
[6] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, Y. SheikhOpenPose: realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017: pp. 1302–1310. https://doi.org/10.1109/CVPR.2017.143.
[7] N. Nakano, T. Sakura, K. Ueda, L. Omura, A. Kimura, Y. Iino, S. Fukashiro, S. Yoshioka, Evaluation of 3D markerless motion capture accuracy using openpose with multiple video cameras, Front. Sports Act. Living 2 (2020) 50, https://doi.org/10.3389/fspor.2020.00050.
[8] Y. Aoyagi, S. Yamada, S. Ueda, C. Iseki, T. Kondo, K. Mori, Y. Kobayashi, T. Fukami, M. Hoshimaru, M. Ishikawa, Y. Ohta, Development of smartphone application for markerless three-dimensional motion capture based on deep learning model, Sensors 22 (2022) 5282, https://doi.org/10.3390/s22145282.

[9] J. Stenum, C. Rossi, R.T. Roemmich, Two-dimensional video-based analysis of human gait using pose estimation, PLOS Comput. Biol. 17 (2021), https://doi.org/10.1371/journal.pcbi.1008935.
[10] T.A.L. Wren, G.E. Gorton, S. Õunpuu, C.A. Tucker, Efficacy of clinical gait analysis: a systematic review, Gait Posture 34 (2011) 149–153, https://doi.org/10.1016/j.gaitpost.2011.03.027.
[11] R. Baker, Gait analysis methods in rehabilitation, J. Neuroeng. Rehabil. 3 (2006) 4, https://doi.org/10.1186/1743-0003-3-4.
[12] M.H. Schwartz, J.P. Trost, R.A. Wervey, Measurement and management of errors in quantitative gait data, Gait Posture 20 (2004) 196–203, https://doi.org/10.1016/j.gaitpost.2003.09.011.
[13] J.L. McGinley, R. Baker, R. Wolfe, M.E. Morris, The reliability of three-dimensional kinematic gait measurements: a systematic review, Gait Posture 29 (2009) 360–369, https://doi.org/10.1016/j.gaitpost.2008.09.003.
[14] S.D. Uhlrich, A. Falisse, J. Kidziński, M. Muccini, A.S. Ko, J.L. Chaudhari, S. L. Hicks, Delp, OpenCap: human movement dynamics from smartphone videos, PLOS Comput. Biol. 19 (2023) e1011462, https://doi.org/10.1371/journal.pcbi.1011462.
[15] J.H. Shin, R. Yu, J.N. Ong, C.Y. Lee, S.H. Jeon, H. Park, H.-J. Kim, J. Lee, B. Jeon, Quantitative gait analysis using a pose-estimation algorithm with a single 2D-video of Parkinson's disease patients, J. Park. Dis. 11 (2021) 1271–1283, https://doi.org/10.3233/JPD-212544.
[16] J. Stenum, M.M. Hsu, A.Y. Pantelyat, R.T. Roemmich, Clinical gait analysis using video-based pose estimation: multiple perspectives, clinical populations, and measuring change, PLOS Digit Health 3 (2024) e0000467, https://doi.org/10.1371/journal.pdig.0000467.
[17] B. Kidziński, J.L. Yang, A. Hicks, S.L. Rajagopal, M.H. Delp, Schwartz, Deep neural networks enable quantitative movement analysis using single-camera videos, Nat. Commun. 11 (2020) 4054, https://doi.org/10.1038/s41467-020-17807-z.
[18] T. Li, J. Chen, C. Hu, Y. Ma, Z. Wu, W. Wan, Y. Huang, F. Jia, C. Gong, S. Wan, L. Li, Automatic timed up-and-go sub-task segmentation for Parkinson's disease patients using video-based activity classification, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (2018) 2189–2199, https://doi.org/10.1109/TNSRE.2018.2875738.
[19] S. Rupprechter, G. Morinan, Y. Peng, T. Foltynie, K. Sibley, R.S. Weil, L.-A. Leyland, F. Baig, F. Morgante, R. Gilron, R. Wilt, P. Starr, R.A. Hauser, J. O'Keeffe, A clinically interpretable computer-vision based method for quantifying gait in parkinson's disease, Sensors 21 (2021), https://doi.org/10.3390/s21165437.
[20] A. Liberati, D.G. Altman, J. Tetzlaff, C. Mulrow, P.C. Gøtzsche, J.P.A. Ioannidis, M. Clarke, P.J. Devereaux, J. Kleijnen, D. Moher, The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration, PLOS Med. 6 (2009) e1000100, https://doi.org/10.1371/journal.pmed.1000100.
[21] P.F. Whiting, QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (2011) 529, https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
[22] K. John, J. Stenum, C.-C. Chiang, M.A. French, C. Kim, J. Manor, M.A. Statton, K. M. Cherry-Allen, R.T. Roemmich, Accuracy of video-based gait analysis using pose estimation during treadmill walking versus overground walking in persons after stroke, Phys. Ther. 104 (2024), https://doi.org/10.1093/ptj/pzad121.
[23] J. Yang, K. Park, Improving gait analysis techniques with markerless pose estimation based on smartphone location, Bioengineering 11 (2024), https://doi.org/10.3390/bioengineering11020141.
[24] J. Lin, Y. Wang, J. Sha, Y. Li, Z. Fan, W. Lei, Y. Yan, Clinical reliability and validity of a video-based markerless gait evaluation method, Front. Pediatr. 11 (2023), https://doi.org/10.3389/fped.2023.1331176.
[25] E.P. Washabaugh, T.A. Shanmugam, R. Ranganathan, C. Krishnan, Comparing the accuracy of open-source pose estimation methods for measuring gait kinematics, Gait Posture 97 (2022) 188–195, https://doi.org/10.1016/j.gaitpost.2022.08.008.
[26] S. Talaa, M.E. Fezazi, A. Jilbab, H.E.Y. Alaoui, Computer vision-based approach for automated monitoring and assessment of gait rehabilitation at home, Int. J. Online Biomed. Eng. 19 (2023) 139–157, https://doi.org/10.3991/ijoe.v19i18.43943.
[27] B. Horsak, A. Eichmann, K. Lauer, K. Prock, P. Krondorfer, T. Siragy, B. Dumphart, Concurrent validity of smartphone-based markerless motion capturing to quantify lower-limb joint kinematics in healthy and pathological gait, J. Biomech. 159 (2023), https://doi.org/10.1016/j.jbiomech.2023.111801.
[28] D. Menychtas, N. Petrou, I. Kansizoglou, E. Giannakou, A. Grekidis, A. Gasteratos, V. Gourgoulis, E. Douda, I. Smilios, M. Michalopoulou, G.C. Sirakoulis, N. Aggelousis, Gait analysis comparison between manual marking, 2D pose estimation algorithms, and 3D marker-based system, Front. Rehabil. Sci. 4 (2023) 1238134, https://doi.org/10.3389/fresc.2023.1238134.
[29] M. Boldo, R. Di Marco, E. Martini, M. Nardon, M. Bertucco, N. Bombieri, On the reliability of single-camera markerless systems for overground gait monitoring, Comput. Biol. Med. 171 (2024), https://doi.org/10.1016/j.compbiomed.2024.108101.
[30] S. Liang, Y. Zhang, Y. Diao, G. Li, G. Zhao, The reliability and validity of gait analysis system using 3D markerless pose estimation algorithms, Front. Bioeng. Biotechnol. 10 (2022) 857975, https://doi.org/10.3389/fbioe.2022.857975.
[31] T. Ino, M. Samukawa, T. Ishida, N. Wada, Y. Koshino, S. Kasahara, H. Tohyama, Validity of AI-based gait analysis for simultaneous measurement of bilateral lower limb kinematics using a single video camera, Sensors 23 (2023), https://doi.org/10.3390/s23249799.
[32] M. Ota, H. Tateuchi, T. Hashiguchi, N. Ichihashi, Verification of validity of gait analysis systems during treadmill walking and running using human pose tracking algorithm, Gait Posture 85 (2021) 290–297, https://doi.org/10.1016/j.gaitpost.2021.02.006.
[33] M. Zago, M. Luzzago, T. Marangoni, M. De Cecco, M. Tarabini, M. Galli, 3D Tracking of human motion using visual skeletonization and stereoscopic vision, Front. Bioeng. Biotechnol. 8 (2020), https://doi.org/10.3389/fbioe.2020.00181.
[34] I. Takeda, A. Yamada, H. Onodera, Artificial intelligence-assisted motion capture for medical applications: a comparative study between markerless and passive marker motion capture, Comput. Methods Biomech. Biomed. Eng. 24 (2021) 864–873, https://doi.org/10.1080/10255842.2020.1856372.
[35] T.K. Koo, M.Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, J. Chiropr. Med. 15 (2016) 155–163, https://doi.org/10.1016/j.jcm.2016.02.012.
[36] R. Votel, N. Li, Next-generation pose detection with MoveNet and TensorFlow.js, 2021. ⟨https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html⟩.
[37] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, M. Grundmann, BlazePose: On-device Real-time Body Pose tracking, 2020. ⟨http://arxiv.org/abs/2006.10204⟩ (accessed June 10, 2024).
[38] W. McNally, K. Vats, A. Wong, J. McPhee, Rethinking keypoint representations: modeling keypoints and poses as objects for multi-person human pose estimation, in: S. Avidan, G. Brostow, M. Cissé, G.M. Farinella, T. Hassner (Eds.), Computer Vision – ECCV 2022, Springer Nature Switzerland, Cham, 2022, pp. 37–54, https://doi.org/10.1007/978-3-031-20068-7_3.
[39] S. Haghayegh, H.-A. Kang, S. Khoshnevis, M.H. Smolensky, K.R. Diller, A comprehensive guideline for Bland–Altman and intra class correlation calculations to properly compare two methods of measurement and interpret findings, Physiol. Meas. 41 (2020) 055012, https://doi.org/10.1088/1361-6579/ab86d6.
[40] R.M. Kanko, E.K. Laende, E.M. Davis, W.S. Selbie, K.J. Deluzio, Concurrent assessment of gait kinematics using marker-based and markerless motion capture, J. Biomech. 127 (2021) 110665, https://doi.org/10.1016/j.jbiomech.2021.110665.
[41] D. Pagnon, M. Domalain, L. Reveret, Pose2Sim: an end-to-end workflow for 3D markerless sports kinematics—part 2: accuracy, Sensors 22 (2022) 2712, https://doi.org/10.3390/s22072712.
[42] S.A.A.N. Bolink, H. Naisas, R. Senden, H. Essers, I.C. Heyligers, K. Meijer, B. Grimm, Validity of an inertial measurement unit to assess pelvic orientation angles during gait, sit–stand transfers and step-up transfers: comparison with an optoelectronic motion capture system, Med. Eng. Phys. 38 (2016) 225–231, https://doi.org/10.1016/j.medengphy.2015.11.009.
[43] M. Al Borno, J. O'Day, V. Ibarra, J. Dunne, A. Seth, A. Habib, C. Ong, J. Hicks, S. Uhlrich, S. Delp, OpenSense: an open-source toolbox for inertial-measurement-unit-based measurement of lower extremity kinematics over long durations, J. Neuroeng. Rehabil. 19 (2022) 22, https://doi.org/10.1186/s12984-022-01001-x.
[44] F. D'Isidoro, C. Brockmann, S.J. Ferguson, Effects of the soft tissue artefact on the hip joint kinematics during unrestricted activities of daily living, J. Biomech. 104 (2020) 109717, https://doi.org/10.1016/j.jbiomech.2020.109717.
[45] D.L. Benoit, M. Damsgaard, M.S. Andersen, Surface marker cluster translation, rotation, scaling and deformation: their contribution to soft tissue artefact and impact on knee joint kinematics, J. Biomech. 48 (2015) 2124–2129, https://doi.org/10.1016/j.jbiomech.2015.02.050.
[46] L. Wade, L. Needham, P. McGuigan, J. Bilzon, Applications and limitations of current markerless motion capture methods for clinical gait biomechanics, PeerJ 10 (2022) e12995, https://doi.org/10.7717/peerj.12995.
[47] M.A. Boswell, J.L. Kidziński, S.D. Hicks, A. Uhlrich, S.L. Falisse, Delp, Smartphone videos of the sit-to-stand test predict osteoarthritis and health outcomes in a nationwide study, Npj Digit. Med. 6 (2023) 32, https://doi.org/10.1038/s41746-023-00775-1.
[48] W. Rahman, M. Hasan, M.S. Islam, T. Olubajo, J. Thaker, A.-R. Abdelkader, P. Yang, H. Paulson, G. Oz, A. Durr, T. Klockgether, T. Ashizawa, R. Investigators, E. Hoque, Auto-Gait: automatic ataxia risk assessment with computer vision from gait task videos, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 7 (2023) 1–19, https://doi.org/10.1145/3580845.
[49] S. Rupprechter, G. Morinan, Y. Peng, T. Foltynie, K. Sibley, R.S. Weil, L.-A. Leyland, F. Baig, F. Morgante, R. Gilron, R. Wilt, P. Starr, R.A. Hauser, J. O'Keeffe, A Clinically interpretable computer-vision based method for quantifying gait in Parkinson's disease, Sensors 21 (2021) 5437, https://doi.org/10.3390/s21165437.
[50] M. Cheriet, V. Dentamaro, M. Hamdan, D. Impedovo, G. Pirlo, Multi-speed transformer network for neurodegenerative disease assessment and activity recognition, Comput. Methods Prog. Biomed. 230 (2023) 107344, https://doi.org/10.1016/j.cmpb.2023.107344.
[51] G.J. L'Italien, E.K. Oikonomou, R. Khera, M.H. Potashman, M.W. Beiner, G.D. H. Maclaine, J.D. Schmahmann, S. Perlman, V. Coric, Video-based kinematic analysis of movement quality in a phase 3 clinical trial of troriluzole in adults with spinocerebellar ataxia: a post hoc analysis, Neurol. Ther. (2024), https://doi.org/10.1007/s40120-024-00625-6.
[52] E. Samkari, M. Arif, M. Alghamdi, M.A. Al Ghamdi, Human pose estimation using deep learning: a systematic literature review, MAKE 5 (2023) 1612–1659, https://doi.org/10.3390/make5040081.
[53] R. Li, Q. Li, S. Yang, X. Zeng, A. Yan, An efficient and accurate 2D human pose estimation method using VTTransPose network, Sci. Rep. 14 (2024) 7608, https://doi.org/10.1038/s41598-024-58175-8.
[54] Y. Chen, Y. Tian, M. He, Monocular human pose estimation: a survey of deep learning-based methods, Comput. Vis. Image Underst. 192 (2020) 102897, https://doi.org/10.1016/j.cviu.2019.102897.

[55] X. Ji, Q. Fang, J. Dong, Q. Shuai, W. Jiang, X. Zhou, A survey on monocular 3D human pose estimation, Virtual Real. Intell. Hardw. 2 (2020) 471–500, https://doi.org/10.1016/j.vrih.2020.04.005.
[56] O. Elharrouss, Y. Akbari, N. Almadeed, S. Al-Maadeed, Backbones-review: feature extractor networks for deep learning and deep reinforcement learning approaches in computer vision, Comput. Sci. Rev. 53 (2024) 100645, https://doi.org/10.1016/j.cosrev.2024.100645.
[57] K. He, X. Zhang, S. Ren, J. SunDeep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, 2016: pp. 770–778. https://doi.org/10.1109/CVPR.2016.90..
[58] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2015. ⟨http://arxiv.org/abs/1409.1556⟩ (accessed July 4, 2024).
[59] K. Sun, B. Xiao, D. Liu, J. WangDeep high-resolution representation learning for human pose estimation. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, 2019: pp. 5686–5696. https://doi.org/10.1109/CVPR.2019.00584..
[60] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, I.-H. YehCSPNet: a new backbone that can enhance learning capability of CNN. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Seattle, WA, USA, 2020: pp. 1571–1580. https://doi.org/10.1109/CVPRW50498.2020.00203..
[61] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. AdamMobileNets: efficient convolutional neural networks for mobile vision applications.2017.
[62] T.-Y. Lin, M. Maire, Microsoft COCO: Common Objects in Context, Springer International Publishing, Cham, 2014, https://doi.org/10.1007/978-3-319-10602-1.
[63] T. Kitamura, H. Teshima, D. Thomas, H. KawasakiRefining openpose with a new sports dataset for robust 2d pose estimation. In: Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), IEEE, Waikoloa, HI, USA, 2022: pp. 672–681. https://doi.org/10.1109/WACVW54805.2022.00074..
[64] J. Archila, A. Manzanera, F. Martínez, A multimodal Parkinson quantification by fusing eye and gait motion patterns, using covariance descriptors, from non-invasive computer vision, Comput. Methods Prog. Biomed. 215 (2022), https://doi.org/10.1016/j.cmpb.2021.106607.