Please cite the Published Version

Additional Information: This is an Open Access article published in Sports Medicine, by Springer.

Data Access Statement: All data generated or analysed during this study are included in this published article and its supplementary information files.

The Role of Multi-Sensor Measurement in the Assessment of Movement Quality: A Systematic Review

T. Alexander Swain[1] · Melitta A. McNarry[1] · Adam W. H. Runacres[2] · Kelly A. Mackintosh[1]

Abstract

Background Movement quality is typically assessed by drawing comparisons against predetermined movement standards. Movements are often discretely scored or labelled against pre-set criteria, though movement quality can also be evaluated using motion-related measurements (e.g., spatio-temporal parameters and kinematic variables). Wearable technology has the potential to measure and assess movement quality and offer valuable, practical feedback.

Objectives A systematic approach was taken to examine the benefits associated with multi-sensor and multiple wearable-device usage, compared with unimodal applications, when assessing movement quality. Consequently, this review considers the additional variables and features that could be obtained through multi-sensor devices for use in movement analyses. Processing methods and applications of the various configurations were also explored.

Methods Articles were included within this review if they were written in English, specifically studied the use of wearable sensors to assess movement quality, and were published between January 2010 and December 2022. Of the 62,635 articles initially identified, 27 papers were included in this review. The quality of included studies was determined using a modified Downs and Black checklist, with 24/27 high quality.

Results Fifteen of the 27 included studies used a classification approach, 11 used a measurement approach, and one used both methods. Accelerometers featured in all 27 studies, in isolation ($n = 5$), with a gyroscope ($n = 9$), or with both a gyroscope and a magnetometer ($n = 13$). Sampling frequencies across all studies ranged from 50 to 200 Hz. The most common classification methods were traditional feature-based classifiers ($n = 5$) and support vector machines (SVM; $n = 5$). Sensor fusion featured in six of the 16 classification studies and nine of the 12 measurement studies, with the Madgwick algorithm most prevalent ($n = 7$).

Conclusions This systematic review highlights the differences between the applications and processing methods associated with the use of unimodal and multi-sensor wearable devices when assessing movement quality. Further, the use of multiple devices appears to increase the feasibility of effectively assessing holistic movements, while multi-sensor devices offer the ability to obtain more output metrics.

1 Introduction

Movement quality is a historically overlooked component of physical activity and exercise monitoring outside of elite sports [1]. Whilst wearable technology has become an integral part of modern lifestyles, wearable movement tracker

✉ Kelly A. Mackintosh
k.mackintosh@swansea.ac.uk

1 Applied Sports, Technology, Exercise and Medicine Research Centre (A-STEM), Swansea University, Swansea, UK

2 Institute of Sport, Manchester Metropolitan University, Manchester, UK

Key Points

Wearable technology has been widely used in research to measure and classify matters relating to movement quality using an array of machine-learning and sensor-fusion methods.

Multiple multimodal sensor devices appear to be more effective than single multimodal sensor devices when assessing holistic movements.

Devices containing multiple sensors offer the ability to obtain more output metrics than those featuring just one sensor.

outputs are almost exclusively focused on movement quantity [2]. However, wearable technology also has the potential to measure and assess movement quality by facilitating specific and contextualised feedback [3]. Indeed, this is advantageous in the reduction of injury risk, given that poor quality movement is widely recognised as a contributor to injury [4–9]. Additionally, better movement quality is recognised to enhance life-expectancy by increasing motivation and confidence to engage in physical activity [10], and skill proficiency is fundamental for optimal athletic performance [11–13]. Furthermore, the development of motor competence across a wide array of motor skills in young people is an essential foundation for overall health throughout the lifecourse [10, 14]. Nonetheless, there remains an unmet demand surrounding the assessment of, and feedback regarding, movement quality, which could benefit greatly from the capabilities of wearable devices [14].

While the definition of movement quality is open to interpretation and likely context-specific, Venek et al. [15] offer a perspicuous, generalised definition of movement quality as "the degree to which replications of the original movements can be performed in comparison to either an expert or professional, or to a defined performance of an exercise" [15]. Further expanding on this definition, in clinical settings, movement quality may be assessed by comparing pathologically influenced movements against healthy controls, or normative data [16, 17]. Consequently, assessing movement quality necessitates a predetermined standard against which comparisons can be made, with discrete scoring or labelling systems often employed to distinguish between good and bad movements, different skill levels, or to highlight specific movement discrepancies [15]. Moreover, movement quality may also be evaluated using motion-based measurements, for example, spatio-temporal parameters and kinematic variables [1], if baseline information exists to draw comparisons against.

Bardid et al. [14] provide an overview of the wide range of options currently available in the context of assessing motor competence in children and adolescents, identifying a spectrum from which appropriate assessment tools can be selected depending on the application and criteria. However, it is postulated that there could be a large degree of applicability of this spectrum for assessing movement quality, beyond the confines of motor development in children and adolescents. Traditionally, movement-quality assessments have frequently been subjectively conducted by experts, such as physiotherapists, teachers, and/or trained assessors [15, 18–20]. However, modern technologies make it feasible to conduct objective assessments [1], and support individuals in the absence of a movement expert [21]. Camera-based technologies such as optical motion capture [22, 23] and depth cameras [22, 24, 25] are commonly used in the assessment of movement quality due to an array of advantages

[22, 24, 25], though such methods are also associated with significant limitations [20, 22, 24, 25]. Wearable technology, however, may provide an affordable, practical, and efficient alternative for assessing movement quality [14, 20, 21, 26, 27], with the potential for automated feedback using numerical data or visualisations [28]. Currently, there are many commercially available wearable devices used to evaluate movement, including unimodal sensors, such as accelerometers, gyroscopes, and magnetometers [29–31], or multimodal sensor devices, for example, an inertial measurement unit (IMU) [32–35]. In addition, multiple multimodal sensors can be utilised, where they are positioned at different anatomical locations to provide additional and integrated outputs [27, 36]. Sensor combinations may enable a broader picture of movements, which would be valuable when assessing movement quality [1, 32].

Congruent with many subjective assessments [37–39], data obtained by technology may be utilised to categorise movement characteristics, using either binary or multi-class classification methods [20]. A binary classification approach is typically utilised to distinguish between a competent or non-competent movement; that is, whether or not an individual has demonstrated movement proficiency based on the criteria of a predetermined standard. A multi-class classification approach, however, adds a degree of specificity by highlighting specific characteristics [20]. The existing literature suggests that the performance of classification algorithms relative to others depends on many factors, such as the movement performed, sensor positioning, and the parameters considered [40–42]. Indeed, the accuracy of any output from a wearable device can be influenced by the hardware utilised, with sampling frequency being a key factor that needs to be optimised to achieve maximum performance [40, 43, 44]. Nevertheless, it is recognised that the use of additional sensors and wearable devices would enable the capture of a broader dataset with more measurable outputs, which would theoretically enable a more in-depth, and possibly accurate, classification by having a greater range of classifier inputs to select from [45]. Sensors are also commonly applied to directly measure specific motion characteristics, including kinematics such as acceleration, velocity, and displacement [46]. However, it is also possible to capitalise on concepts such as sensor fusion to estimate orientations and angles when multiple types of sensors are available within a device [45, 47, 48].

Previous reviews have provided insights into the potential use of wearable technology to widely assess movement quality [15, 20, 21, 49]. Most recently, a scoping review was conducted that provided a timely update on the array of technology-based measurement methods available for assessing movement quality, though the review only sought to provide a broad overview of technology usage in sport over a 5-year period [15]. Moreover, there is a lack

of consideration in abovementioned reviews of the potential benefits of concurrent sensor usage, either within the same multimodal sensor unit, or using multiple devices, when compared to unimodal alternatives [15, 20, 21, 49]. The current systematic review, therefore, sought to highlight any additional benefits that could be deduced when concurrently applying multi-sensor devices, and indeed multiple wearable devices, for the assessment of movement quality, contextualised around sporting and clinical applications. Consequently, this review also investigated the additional variables, and indeed features, that could be obtained through multi-sensor devices compared to unimodal sensors for use in movement analyses. Finally, this systematic review aimed to distinguish between the processing methods and applications of multi-sensor wearable devices in comparison to unimodal sensors when assessing movement quality.

2 Methods

The protocol was developed to implement a systematic approach in accordance with guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [50, 51]. A Population Intervention Comparison Outcome (PICO) approach to the development of the systematic review framework was also employed. Details of the protocol for this systematic review were registered on PROSPERO (ID: CRD42020222587).

2.1 Search Strategy

An initial electronic literature search was completed between November and December 2020, with additional searches conducted to incorporate any additional relevant publications up to December 2022. The searches sought to identify published materials indexed in the following five databases: MEDLINE, ACM Digital Library, IEEE Xplore, SPORTDiscus, and Scopus. Initial keywords 'movement', 'quality', 'wearable' and 'human' were identified as suitable group headings for an expanded search of other related search terms and synonyms, with a Boolean search strategy implemented thereafter (Table 1; Online Supplemental Resource (OSR) 1). The search strategy utilised keywords in lieu of subject headings to broaden the search. Search terms were sought from within the title, abstract and listed keywords for each publication.

2.2 Literature Screening and Selection

Articles retrieved from the electronic databases were initially stored using Mendeley referencing software (Mendeley Desktop, Version 1.19.8) before uploading to Rayyan [52] for screening. The PRISMA flowchart [50] was utilised to record the process of screening and selection. Duplicates arising from the use of multiple databases were identified within Rayyan and removed. Initially, titles and abstracts were screened for full-text review according to the inclusion and exclusion criteria (Table 2). The screening process was conducted by two authors (TAS and AWHR) and was blinded [52]; that is, each author screened the articles independently, with discrepancies revealed on conclusion of the screening phase. Inter-rater agreement was assessed at each stage of the screening process using kappa scores [53]. All articles were screened by the first author (TAS), while another researcher (AWHR) screened 10% [54, 55]. It is reasonable when undertaking large reviews for two reviewers to initially screen a small percentage of studies and discuss variation in the interpretation of the inclusion criteria [55]. In doing so, this aids consistency when the remaining studies are screened by a single reviewer for inclusion within the full-text screening phase [55]. A kappa score of 0.955, an

Table 1 Boolean search strategy

General	Specific search terms
Movement	(movement* OR motion OR "human mechanics" OR kinematics OR biomechanics OR locomot* OR "motor skill*" OR gait OR ambulat*)
	AND
Quality	(quality OR proficien* OR competen* OR performance OR abilit*)
	AND
Wearable	(wearable* OR sensor* OR multi-sensor* OR multisensor* OR "inertial measurement unit*" OR IMU$ OR acceleromet* OR gyroscop* OR magnetomet* OR detector*)
	AND
Human	human* OR men OR women OR child* OR people OR adolescent$ OR teenagers OR athlete*

An asterisk (*) expanded the search to include all terms beginning with the letters preceding the asterisk. A dollar sign ($) expanded the search to include a single additional letter in the position of the dollar sign. The dollar sign was substituted with a question mark (?) when searching ACM Digital Library due to the criteria for this database

Table 2 Study inclusion and exclusion criteria

Inclusion criteria

Wearable sensors were used for the purpose of assessing movement quality in medical and sporting applications

The wearable sensors were used to assess fine or gross motor skills

Human participants irrespective of age

The wearable devices were attached to individuals directly, or to manually operated equipment (i.e. sports equipment)

Articles were published between January 2010 and December 2022

Peer-reviewed articles published in the English language

Experimental and analytical observational studies

Exclusion criteria

Descriptive and qualitative studies

Systematic reviews and literature reviews

Non-peer-reviewed literature

Other unsuitable resources such as conference presentations, expert opinion, and grey literature

Studies that feature particularly varied or irregular movements, or sequential combination movement patterns (i.e., multiple discrete movement patterns performed transitionally in series)

Studies where wearable outputs were not assessed

almost perfect agreement [53], was observed during the title and abstract screening, after which the discrepancies among the dual-screened studies were resolved through discussion between the two reviewers. All full texts of the remaining articles were then screened against the inclusion and exclusion criteria by the same two reviewers, again independently under blind conditions. A weak agreement was initially observed, generating a kappa score of 0.491. However, following discussions to clarify the inclusion criteria, any conflicts were discussed until a consensus was reached. Where the two reviewers could not agree ($n = 2$), the discrepancies were discussed with a third independent reviewer (KAM) until a consensus was reached, subsequently resulting in a kappa score of 1.

2.3 Data Extraction

Following screening, data were extracted from the included full texts by the lead author (TAS) and tabulated within a customised data extraction form. The data extraction form was subsequently reviewed by another author (MAM); where discrepancies were identified, the data extraction form was revised until both authors reached a consensus.

2.4 Quality Assessment

Two authors (TAS and AWHR) evaluated the quality of the included studies to determine the risk of bias. The primary author (TAS) assessed the quality of all studies, while the other author (AWHR) assessed the quality of 14 of the 27

included studies, approximately 50%, in accord with Pai and McGrady [56, 57]. A modified version of the Downs and Black [58] checklist was selected as it can be used to assess the methodological quality of both randomised controlled trials and non-randomised studies. Specifically, the modifications made to the Downs and Black checklist [58], which were based on other systematic reviews synthesising the use of wearable technology [21, 49], ensured specificity and, therefore, relevance to the included studies. The criteria were accompanied by a rating system developed to categorise study quality as low ($\leq 33.3\%$), moderate ($33.4–66.7\%$), and high ($\geq 66.8\%$; OSR 2). The limits of each scoring category were adopted from other systematic reviews focused on the use of wearable technology to measure and assess movement [21, 49]. Eighteen items were included in the checklist, with each item rated between zero and two ($0 =$ not present, $1 =$ limited detail and $2 =$ good detail). Inter-rater reliability was calculated using kappa scores, with 0.8 identified as the minimum acceptable inter-rater agreement [53]. Initially, the kappa score for risk of bias was 0.273, indicative of fair agreement [53]. Discrepancies centred around study design, eligibility criteria, and the reliability of equipment, primarily due to each author's interpretation of the checklist questions. The two authors discussed each point until a consensus was reached, resulting in a kappa score of 1. Subsequently, the remaining studies were reassessed based on the agreed interpretation of each quality assessment checklist item.

Given the broad inclusion criteria and subsequent methodological differences in the included studies, a narrative synthesis was conducted. The methodological approach is covered, with particular focus on the identification of sensor features, the application of sensors and the obtained data, and the techniques by which movement quality was assessed. Data processing and analysis methods, including sensor-fusion algorithms, machine-learning techniques and biological modelling, were of particular interest. The findings of the included studies were assessed, specifically focusing on the comparison between unimodal and independent uses of sensors and devices, and more systemic approaches, where the data obtained from multiple sensors and devices were concurrently utilised and integrated.

3 Results

A total of 62,635 titles were obtained across the five databases, with 58,231 titles remaining after the removal of duplicates. Following the screening of 80 full-texts, 27 articles were included in the final review (Fig. 1). For the quality assessment, 24 of the included studies were deemed to be of high quality (score > 66.8%), with three identified as moderate quality (score 33.4–66.7%; Table 3).

3.1 Device and Sensor Specifications and Applications

An array of devices was utilised in the included studies, with little commonality (Table 4). The sensor configurations and their applications within each study are detailed in Tables 5 and 6, respectively. All included studies utilised a wearable device that included an accelerometer, either in isolation ($n = 5$), in combination with a gyroscope ($n = 9$), or with both a gyroscope and a magnetometer ($n = 13$). Notably, however, eight of the 27 studies did not utilise the full sensory capabilities of the device within their study (Table 4) [62, 67–69, 76, 81, 83, 84]. There was a similar degree of variation in the sampling frequencies utilised (50–200 Hz; Table 4), which were reported in all but one study [59].

The wearable devices were positioned across a broad range of anatomical locations (Tables 5 and 6), most commonly on the lower back ($n = 15$; [60–62, 64, 67, 68, 70, 72, 73, 75, 77, 78, 80, 81, 83]) and distal leg segments, encompassing the shanks and ankles ($n = 14$; [59, 61, 66, 68, 69, 72, 74, 76–78, 81, 83–85]). Other anatomical locations for the sensors were the wrists or forearms ($n = 8$; [61, 63, 67, 69, 71, 72, 79, 81]), thighs ($n = 8$; [59, 68, 72, 77, 78, 80, 81, 83]), chest ($n = 6$; [63, 65, 67, 71, 81, 85]), feet ($n = 5$; [72, 74, 80, 81, 84]), mid back ($n = 4$; [60, 72, 75, 81]), upper arms ($n = 4$; [63, 71, 72, 81]), pelvis ($n = 1$; [65]), hip ($n = 1$; [83]), hand ($n = 1$; [81]), and head ($n = 1$; [81]). Additionally, one study positioned a single wearable device on a piece of sports equipment, specifically a table tennis racket [82]. The number of wearable devices used ranged from one to 17, with 25 of the 27 studies featuring seven or fewer devices. Moreover, it was indicated by Liu et al. [72] that not all 12 devices were utilised throughout their entire study; only the upper body was considered in the measurement of joint angles.

Sensor fusion was a common feature across the included studies, with 14 of the 27 studies capitalising on such an approach. The Madgwick algorithm [48] was employed in seven studies [59, 63, 72, 75, 77–79], while a Kalman filter [86] was used in five studies [60, 66, 68, 80, 82], and complementary filters [87, 88] were used in two studies [71, 74]. When using wearables to measure kinematics, there were exceptions to the use of sensor fusion, with Del Din et al. [64] utilising a single accelerometer, Mitternacht et al. [76] using both accelerometers and gyroscopes, and Tulipani et al. [83] employing gyroscopes exclusively. In the studies by Del Din et al. [64] and Tulipani et al. [83], sensor fusion was not possible due to the omission of other sensors. Rather, Del Din et al. [64] applied the inverted pendulum model [89] to measure step length, while Tulipani et al. [83] integrated the angular velocity obtained from the gyroscope using a proprietary algorithm to calculate the angular displacement of body segments. Mitternacht et al. [76] also opted to utilise an integration method for the movements considered within this systematic review, integrating the angular velocity to determine angles, and double integrating acceleration to calculate linear position. The influence of integration drift [32] was reduced by dividing longer time series data into shorter segments [76]. Where sensor fusion was used to classify movement characteristics [66, 68, 72, 77, 78, 82], the orientation data were utilised to expand the available range of features for implementation within the classification algorithms.

3.2 Descriptive Aspects of Reviewed Studies

Of the 27 studies included in the systematic review, 16 featured methods that were used to classify movement qualities or abnormalities (Table 5), and 12 used methods to measure motion-based characteristics (Table 6) to assess movement quality, typically through the measurement of joint angles or segment-rotation angles [59, 60, 63, 71, 72, 74–76, 79–81, 83]. Notably, one study directly measured motion to assess movement quality while also employing classification methods, and was therefore included in both study groupings [72]. Four of the 27 studies used device-based measures to assess the movement of healthy children using wearables [61, 67, 69, 73], all of which utilised classification methods for assessing motor competence when performing fundamental movement skills (FMS). Of the remaining studies, 19 featured healthy adult samples [60, 62, 63, 65, 66, 68, 70, 74–85], three included both healthy participants and participants with a pathology for comparative purposes [59, 64, 71], while one study did not report participant details [72]. Additionally, 11 of the 27 studies were conducted in a clinical setting for medical applications [59, 60, 64, 66, 68, 71, 74, 85], with the remainder focused on sports performance or sports-injury prevention; all included articles were feasibility studies. Sample sizes ranged from two to 77 participants, though 20 studies featured small sample sizes of 20 or fewer participants [59, 60, 62, 63, 65, 66, 68–72, 74, 76, 79–85].

3.3 Classification and Validation Methods Used to Assess Movement Quality

Ten of the 16 classification studies incorporated a binary-classification approach based on the data obtained from the wearable sensors [61, 62, 65, 68–70, 72, 77, 78, 85], while only nine studies applied a multi-class classification [62, 66–68, 73, 77, 78, 84]. Four studies employed and compared both binary- and multi-class methods [62, 68, 77, 78]. In addition, Tabrizi et al. [82] employed multivariate regression models, which, by definition, use continuous
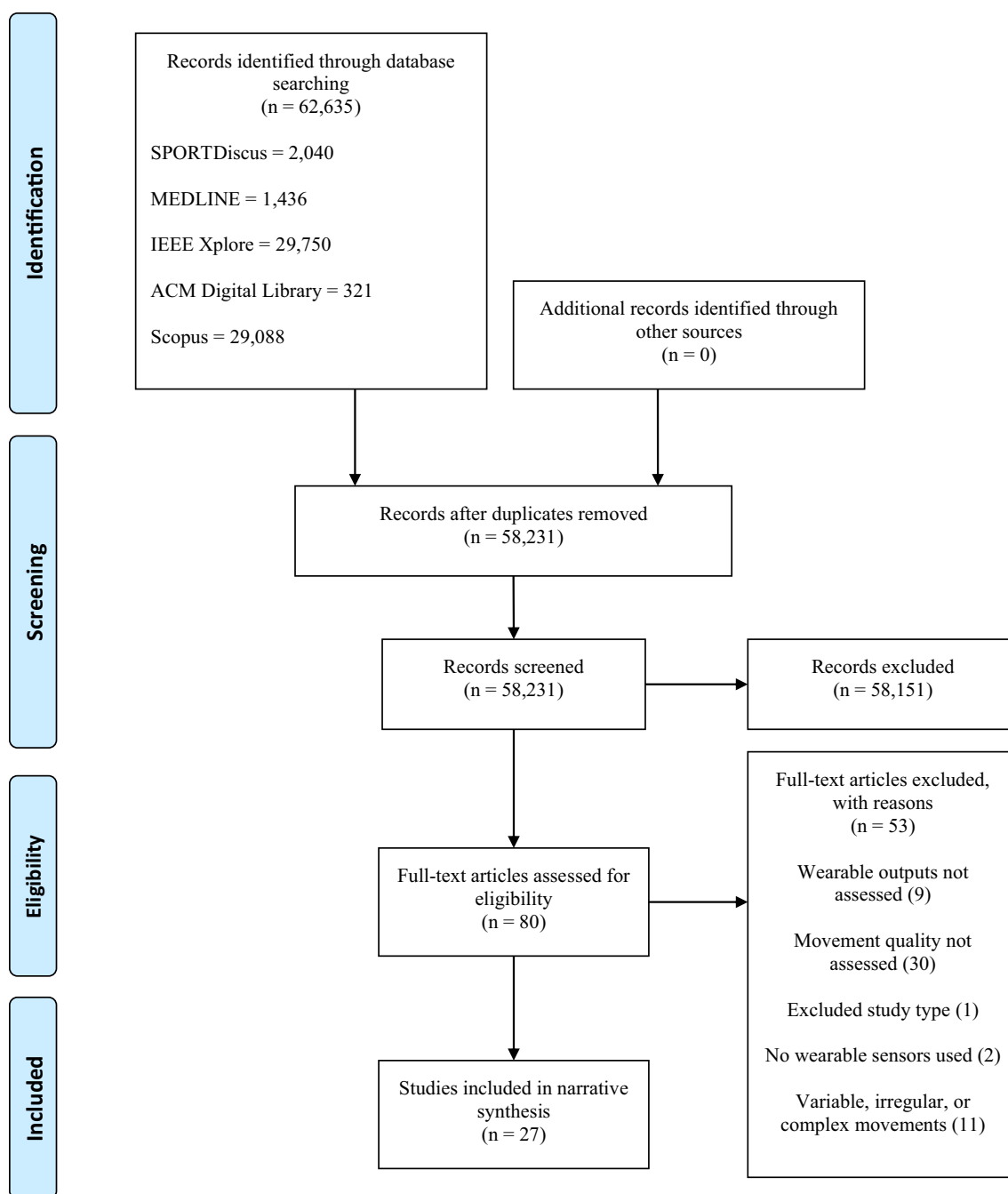
Fig. 1 PRISMA flowchart [50]

and numeric outputs rather than featuring distinct classes. Given the comparable characteristics between classification and regression models, they have been categorised together for this review. Specifically, Tabrizi et al. [82] used conventional machine-learning methods, namely support vector regression (SVR), and two deep-learning methods, convolutional neural networks (CNN) and long short-term memory (LSTM), which were used to generate movement quality scores using set criteria [82]. For the true classification

studies, where accuracies were presented as percentages, binary skill-classification accuracy ranged from 69 to 100% (Table 5) [61, 62, 68–70, 72, 77, 78, 85]. Ghasemzadeh and Jafari [65] also applied a binary-classification approach, though the results were defined in terms of the percentage error of the wearable system relative to analysed video footage (3.4%). Furthermore, while Ghobadi and Esfahani [66] utilised a multi-class approach to distinguish different locomotor activities, there appears to be an intra-classification

Table 3 Quality assessment for included studies

Study	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total score (out of 36)	Percentage score (%)	Quality category
Ahmadi et al. (2015) [59]	2	2	1	1	1	0	2	1	2	2	2	2	2	1	1	2	2	0	26	72.2	High
Beange et al. (2019) [60]	2	2	2	1	0	0	2	2	2	2	1	2	2	2	2	2	2	2	30	83.3	High
Bisi et al. (2017) [61]	2	2	2	1	1	0	2	2	2	2	2	2	2	2	1	2	2	2	31	86.1	High
Caporaso and Grazioso (2020) [62]	2	2	2	2	1	0	2	2	2	2	0	2	2	2	2	2	2	0	29	80.6	High
Cortesi et al. (2017) [63]	2	2	2	1	1	0	2	1	2	2	2	2	1	2	2	2	2	2	30	83.3	High
Del Din et al. (2016) [64]	2	2	2	1	1	0	2	1	2	2	2	2	2	1	2	2	2	2	30	83.3	High
Ghasemzadeh and Jafari (2011) [65]	2	1	1	1	0	0	2	2	2	1	0	2	2	1	0	1	2	0	20	55.6	Moderate
Ghobadi and Esfahani (2017) [66]	2	2	1	1	0	0	2	1	2	2	1	2	2	1	2	2	2	1	26	72.2	High
Grimpampi et al. (2016) [67]	2	2	2	1	1	0	2	2	2	2	2	2	2	1	2	2	2	1	30	83.3	High
Kianifar et al. (2017) [68]	2	2	2	2	0	0	2	1	2	2	2	2	2	2	2	2	2	2	31	86.1	High
Lander et al. (2020) [69]	2	2	2	2	2	0	2	2	2	1	2	2	2	2	2	2	2	2	33	91.7	High
Lee et al. (2013) [70]	2	2	2	0	1	0	2	2	2	2	0	2	2	2	2	2	2	2	29	80.6	High
Lin et al. (2021) [71]	1	2	2	2	1	0	2	2	2	2	0	2	2	2	2	2	2	2	30	83.3	High
Liu et al. (2020) [72]	2	2	2	0	0	0	2	1	2	1	0	2	2	2	2	2	2	2	26	72.2	High
Masci et al. (2013) [73]	2	2	2	2	1	0	2	2	2	2	2	2	2	2	2	2	2	0	31	86.1	High
Meng et al. (2019) [74]	2	2	1	0	0	0	2	1	2	1	0	2	2	2	2	2	2	1	24	66.7	Moderate
Michaud et al. (2021) [75]	2	2	2	1	1	0	2	1	2	2	2	2	2	2	2	2	2	2	31	86.1	High
Mitternacht et al. (2022) [76]	2	2	2	1	1	2	2	2	2	2	0	2	2	2	2	2	2	2	32	88.9	High
O'Reilly et al. (2017a) [77]	2	2	2	2	1	0	2	2	2	2	2	2	2	1	2	2	2	2	32	88.9	High
O'Reilly et al. (2017b) [78]	2	2	2	2	1	0	2	2	2	2	2	2	2	1	2	2	2	2	32	88.9	High
Shepherd et al. (2017) [79]	2	2	2	0	0	0	2	1	2	2	2	2	2	2	2	2	2	0	27	75.0	High
Shuai et al. (2022) [80]	2	2	2	2	1	0	2	1	2	2	2	2	2	2	2	2	2	2	32	88.9	High
Spilz & Munz (2022) [81]	2	2	2	1	1	0	2	1	2	2	2	2	2	1	2	2	2	1	30	83.3	High
Tabrizi et al. (2021) [82]	2	2	2	1	1	0	2	1	2	2	2	2	2	2	2	2	2	2	30	83.3	High
Tulipani et al. (2018) [83]	2	1	2	2	1	0	2	1	2	2	2	2	2	2	2	2	2	2	31	86.1	High
Xu et al. (2015) [84]	1	1	1	1	0	0	2	2	2	1	0	2	2	1	2	1	2	1	22	61.1	Moderate
Zhang et al. (2014) [85]	2	0	2	2	0	0	2	0	2	2	2	2	2	2	0	2	2	2	26	72.2	High

Quality assessment criteria: 1. Were the research objectives or aims clearly stated? 2. Was the study design clearly described? 3. Was the study population/sample adequately described? 4. Were the eligibility criteria specified? 5. Was the sampling methodology appropriately described? 6. Was the sample size used justified? 7. Did the method description enable accurate replication of the measurement procedures? 8. Was the assessed movement(s) sufficiently standardised? 9. Was the equipment design and set up clearly described? 10. Were sensor locations accurately and clearly described? 11. Was the sensor attachment method clearly described? 12. Was the signal/data handling described? 13. Were the main outcomes measured and the related calculations (if applicable) clearly described? 14. Was the system compared to an acknowledged gold standard or other acceptable reference standard? 15. Were measures of reliability/accuracy of the equipment used relative to the reference standard system for the intended outcome reported? 16. Were the main findings of the study stated? 17. Were the statistical tests appropriate? 18. Were limitations of the study clearly described? Criteria Scores: 0 = no evidence, 1 = limited evidence, 2 = criteria met

Table 4 Wearable device systems

Study	Device name	Components				Sampling frequency
		Accelerometer	Gyroscope	Magnetometer	Other	
Ahmadi et al. (2015) [59]	Novel system developed	X	X			–
Beange et al. (2019) [60]	MetaMotionR	X	X	X		100 Hz
Bisi et al. (2017) [61]	Opal	X	X	X		128 Hz
Caporaso and Grazioso (2020) [62]	G-Sensor 2	X	X[a]			200 Hz
Cortesi et al. (2019) [63]	Opal	X	X	X		128 Hz
Del Din et al. (2016) [64]	AX3	X				50/100 Hz[b]
Ghasemzadeh and Jafari (2011) [65]	TelosB	X	X			50 Hz
Ghobadi and Esfahani (2017) [66]	x-IMU	X	X	X		50 Hz
Grimpampi et al. (2016) [67]	Opal	X	X	X[a]		128 Hz
Kianifar et al.(2017) [68]	3-Space (specific model not stated)	X	X	X[a]		90 ± 10 Hz, resampled to 200 Hz
Lander et al. (2020) [69]	MVN Awinda wireless motion capture suit	X	X[a]	X[a]	X[a]	50 Hz
Lee et al. (2013) [70]	Derivative of an inertial measurement system developed by [90]	X				100 Hz
Lin et al. (2021) [71]	Novel system developed	X	X	X		75 Hz
Liu et al. (2020) [72]	Novel inertial measurement system developed by Dalian University of Technology	X	X	X		148 Hz
Masci et al. (2013) [73]	FreeSense	X	X			100 Hz
Meng et al. (2019) [74]	Trigno (specific model not stated)	X	X			51.2 Hz
Michaud et al. (2021) [75]	STT-IWS	X	X	X		100 Hz
Mitternacht et al. (2022) [76]	InvenSense ICM-20948	X	X	X[a]		200 Hz
O'Reilly et al. (2017a) [77]	Shimmer3 IMU	X	X	X		51.2 Hz
O'Reilly et al. (2017b) [78]	Shimmer3 IMU	X	X	X		51.2 Hz
Shepherd et al. (2017) [79]	SABELSense	X	X	X		100 Hz
Shuai et al. (2022) [80]	Perception Neuron	X	X	X		100 Hz
Spilz & Munz (2022) [81]	Shimmer3 IMU	X	X	X	X[a]	120 Hz
Tabrizi et al. (2021) [82]	BNO055	X	X	X		70 Hz
Tulipani et al. (2018) [83]	Opal	X	X	X[a]		128 Hz
Xu et al. (2015) [84]	InvenSense MotionFit	X	X[a]	X[a]		200 Hz
Zhang et al. (2014) [85]	TEMPO 3.1 [91]	X	X			100 Hz

[a]Sensor was present but not utilised as part of the study

[b]50 Hz initially, then device upgraded during the study to accommodate 100 Hz, which was resampled to 50 Hz

binary approach used to distinguish healthy and erroneous gait patterns, with 99% accuracy reported for the detection of abnormalities. Where a multi-class approach was otherwise used, accuracies ranged from 53 to 99% [62, 68, 73, 77, 78, 84]. Grimpampi et al. [67], and Spilz and Muntz [81] also used multi-class classification approaches, though the findings were reported based on the significance of differences between signal features in the former, and in terms of F1-scores [93] in the latter.

A range of classification methods were applied within the included studies (see Table 5). Most common were feature-based classification methods, an overarching term used to encompass traditional feature-learning methods, utilised in five of the 16 studies [61, 62, 67, 70, 73], and SVM, also used in five studies [66, 68, 72, 84, 85]. SVR, derived from traditional SVM modelling, was also applied in Qiu et al. [82]. Other classifiers used throughout the included studies were k-nearest neighbour (KNN) [68], naïve Bayes (NB)

Table 5 Included studies using a classification method to assess movement quality

Study	Participants	Utilised sensors and placements	No. devices	Application	Categorisation	Classification method, input variables and features	Validation method	Findings
Bisi et al. (2017) [61]	45 healthy children (28 boys; 6–10 years)	Tri-axial accel+gyro+mag (low back, ankle L&R, wrist L&R)	5	Motor competence of FMS	Sequential binary technique classifiers (3–4 criteria per movement)	Feature-based classification. Signal features extracted from raw accel, gyro, and mag time-series data	Real-world validation by three experts analysing video footage	Performance criteria classification accuracy = 77–100% Overall skill classification accuracy = 87–96%
Caporaso and Grazioso (2020) [62]	9 healthy adults (7 male; 25.3±4.7 years)	Tri-axial accel (low back)	1	Race walking infringement detection	Binary infringement classification Three-level infringement classification Fuzzy classification	Feature-based classification. Signal features extracted from raw accel time-series data	Real-world validation using video footage from a high-speed camera	Binary infringement classification accuracy = 87% Three-level infringement classification accuracy = 63% Fuzzy classification accuracy = 92%
Ghasemzadeh and Jafari (2011) [65]	3 healthy adults (3 males; 25–35 years)	Tri-axial accel+bi-axial gyro (chest, hip, wrist)	3	Coordination analysis of a baseball swing	Transcript-based coordination analysis to distinguish between proper and improper technique	Statistical feature-based classification utilising k-means clustering Signal features extracted from raw accel and gyro time-series data	Real-world validation using video footage from two webcams which was synchronised with the sensors	The overall error of the sensor-based approach over all categories was 101 ms, equal to 3.4% of the total length of the template (3 s) when compared with the video-based analysis
Ghobadi and Esfahani (2017) [66]	Part 1: 10 healthy adults (sex and age not specified) Part 2: 2 healthy adults from Part 1 sample (1 male; age not specified)	Tri-axial accel+gyro+mag (ankle)	1	Gait abnormality monitoring	Multi-label activity classification	SVM Signal features extracted from raw accel, gyro, and mag time-series data and from orientations obtained through Kalman-based sensor fusion	KF-CV	Detection of abnormal walking accuracy = 99%

Table 5 (continued)

Study	Participants	Utilised sensors and placements	No. devices	Application	Categorisation	Classification method, input variables and features	Validation method	Findings
Grimpampi et al. (2016) [67]	58 healthy children (30 boys; 8 ± 2 years)	Tri-axial accel + gyro (low back, chest, wrist)	3	Motor competence of overarm throwing	Multi-label development level classification	Feature-based classification Signal features extracted from raw accel and gyro time-series data, and derived kinematic parameters	Real-world validation by one expert analysing video footage (camera details not provided)	Between three development levels (DLTC), a significant difference ($p = 0.001$) was achieved between $DLTC_1$ and $DLTC_3$, and $DLTC_2$ and $DLTC_3$, using temporal parameters for the cocking phase of an overarm throw. No significant difference was observed using other temporal parameters. Trunk max acc_{AP} in $DLTC_3$ was significantly higher ($p < 0.0005$) than in $DLTC_1$ and $DLTC_2$, with no significant difference between $DLTC_1$ and $DLTC_2$. A significant difference ($p < 0.0005$) between $DLTC_1$ and $DLTC_2$ for ROM_{Pitch} was also revealed

Table 5 (continued)

Study	Participants	Utilised sensors and placements	No. devices	Application	Categorisation	Classification method, input variables and features	Validation method	Findings
Kianifar et al.(2017) [68]	14 healthy adults (7 males; 30.8 ± 5.5 years)	Tri-axial accel + gyro (low back, thigh, shank)	1 and 3	Technical performance the single-leg squat in a clinical setting	Binary technique classification Three-level technique classification	SVM KNN NB Signal features extracted from raw accel, gyro, and mag time-series data and from joint angles obtained through Kalman-based sensor fusion	KF-CV LOSOCV	Binary classification (KF-CV): 3 IMU–accuracy = 96%, single IMU (ankle)–accuracy: 90% Three-level classification (KF-CV): 3 IMU–accuracy = 65%, single IMU (ankle)–accuracy: 64% Binary classification (LOSOCV): 3 IMU–accuracy = 90%, single IMU (ankle)–accuracy: 69% Three-level classification (LOSOCV-CV): 3 IMU–accuracy = 60%, single IMU (ankle)–accuracy: 50%
Lander et al. (2020) [69]	14 healthy children (9 boys; 7–12 years)	Tri-axial accel (ankle L&R, wrist L&R)	4	Motor competence of FMS	Sequential binary technique classifiers (3–4 criteria per movement)	k-means clustering Signal features extracted from raw accel time- and frequency-domain data	Real-world validation by two experts analysing video footage and data from an IMU motion capture suit	Performance criteria classification accuracy = 72–100% Overall skill classification accuracy = 77–100%
Lee et al. (2013) [70]	7 healthy adults (5 males; 25 ± 7.3 years)	Tri-axial accel (low back)	1	Race walking infringement detection	Binary infringement classification	Feature-based classification Signal features extracted from raw accel time-series data	Real-world validation using video footage	Binary infringement classification accuracy = 91%

Table 5 (continued)

Study	Participants	Utilised sensors and placements	No. devices	Application	Categorisation	Classification method, input variables and features	Validation method	Findings
Liu et al. (2020) [72]	6 (sex, age and medical status not specified)	Tri-axial accel + gyro + mag (upper arm L&R, forearm L&R, low back, mid back, thigh L&R, shank L&R, foot L&R)	12	Canoeing proficiency level	Binary proficiency level classification	SVM Logistic regression Decision tree XGBoost Signal features extracted from raw accel, gyro, and mag time- and frequency-domain data and from joint angles obtained through Madgwick [48] sensor fusion	KF-CV	SVM general features accuracy = 100%, selected features accuracy = 97% Logistic regression general features accuracy = 99%, selected features accuracy = 96% Decision tree general features accuracy = 94%, selected features accuracy = 94% XGBoost general features accuracy = 100%, selected features accuracy = 99%
Masci et al. (2013) [73]	54 healthy children (sex not specified; 5 ± 3 years)	Tri-axial accel + gyro (low back)	1	Motor competence of running based on arm actions	Multi-label technique classification	Feature-based classification Signal features extracted from raw accel and gyro time- and frequency-domain data, and derived kinematic parameters	LOOCV	DLAA detection accuracy = 67–87%

Table 5 (continued)

Study	Participants	Utilised sensors and placements	No. devices	Application	Categorisation	Classification method, input variables and features	Validation method	Findings
O'Reilly et al. (2017a) [77]	80 healthy adults (57 males; 24.68 ± 4.91 years)	Tri-axial accel + gyro + mag (low back, thigh L&R, shank L&R)	1–5	Technical performance of a barbell deadlift	Binary technique classification Multi-label technique classification	Random forests Signal features extracted from raw accel, gyro, and mag time- and frequency-domain data and from joint angles obtained through Madgwick [48] sensor fusion	Global: LOSOCV Personalised LOOCV	Binary classification with natural deviations using a global model (LOSOCV): 5 IMU–accuracy = 73%, single IMU (right thigh)–accuracy: 71% Binary classification with natural deviations using personalised model (LOOCV): 5 IMU–accuracy = 84%, single IMU (right thigh)–accuracy = 82% Multi-class classification of natural deviations using global model (LOSOCV): 5 IMU–accuracy = 54%, single IMU (right shank)–accuracy = 53% Multi-class classification of natural deviations using personalised model (LOOCV): 5 IMU–accuracy = 78%, lumbar IMU–accuracy = 75%

Table 5 (continued)

Study	Participants	Utilised sensors and placements	No. devices	Application	Categorisation	Classification method, input variables and features	Validation method	Findings
O'Reilly et al. (2017b) [78]	77 healthy adults (55 males; 22.63 ± 4.87 years)	Tri-axial accel + gyro + mag (low back, thigh L&R, shank L&R)	1–5	Technical performance of a body-weight squat	Binary technique classification Multi-label technique classification	Random forests Signal features extracted from raw accel, gyro, and mag time- and frequency-domain data and from joint angles obtained through Madgwick [48] sensor fusion	LOSOCV	Binary classification (LOSOCV): 5 IMU–accuracy = 98%, single IMU (left thigh)–accuracy: 98% Multi-class classification (LOSOCV): 5 IMU–accuracy = 80%, single IMU (right shank)–accuracy = 73%
Spilz & Munz (2022) [81]	17 healthy adults (9 males; 24–62 years)	Tri-axial accel + gyro + mag (forehead, chest, upper arm L&R, forearm L&R, hand L&R, mid back [scapulae] L&R, low back, thigh L&R, shank L&R, foot L&R)	17	Technical performance of Functional Movement Screen [92] exercises	Three-level proficiency classification	CNN-LSTM layered neural network Features determined by CNN layer of the neural network	KF-CV LOSOCV	Mean of the macro F1-scores [93] for known participants: Training data = 0.58–0.95, validation data = 0.57–0.95, test data = 0.55–0.91 Mean weighted F1-scores [93] for unknown participants: Training data = 0.81–0.98, validation data = 0.79–0.95, test data = 0.15–0.49

Table 5 (continued)

Study	Participants	Utilised sensors and placements	No. devices	Application	Categorisation	Classification method, input variables and features	Validation method	Findings
Tabrizi et al. (2021) [82]	16 healthy adults (mixed sex; 19–38 years)	Tri-axial accel + gyro + mag (racket blade centre)	1	Technical performance of table tennis forehand strokes	Multivariate regression[a]	SVR LSTM CNN Signal features extracted from raw accel, gyro, and mag time-series data and from orientations obtained through Madgwick [48] sensor fusion	KF-CV	SVR: Weighted average of R^2 = 0.962 LSTM: Weighted average of R^2 = 0.997 CNN: Weighted average of R^2 = 0.958 The weighted average of RMSE showed the LSTM models to have lower estimation error than SVR (4.24%) and CNN (33.79%)
Xu et al. (2015) [84]	6 healthy adults (4 males; age not specified)	Tri-axial accel (foot)	1	Cycling pedal profiling	Multi-label technique classification	SVM Signal features extracted from raw accel time-series data	KF-CV	Pedal profile classification accuracy = 84–99%
Zhang et al. (2014) [85]	17 healthy adults (9 males; 29±11 years)	Tri-axial accel + bi-axial gyro (x nd y axes) + uni-axial gyro (z axis) (shank, chest)	2	Gait abnormality monitoring	Binary fatigue status classification	SVM Signal features extracted from raw accel and gyro time- and frequency-domain data, and derived kinematic parameters	KF-CV	Intra-individual general features accuracy = 97% Inter-individual general features accuracy = 90% Inter-individual selected features accuracy = 93% Inter-individual selected features accuracy = 88%

Findings are reported to the degree of accuracy provided in the original study

accel accelerometer, *CNN* convolutional neural network, *DLAA* development level (arm action), *DLTC* development level (trunk component), *FMS* fundamental movement skill, *gyro* gyroscope, *IMU* inertial measurement unit, *KF-CF* k-fold cross-validation, *KNN* k nearest neighbourhood, *L* left, *LOOCV* leave-one-out cross-validation, *LOSOCV* leave-one-subject-out cross-validation, *LSTM* long short-term memory, *mag* magnetometer, *NB* naïve Bayes, *R* right, R^2 adjusted coefficient of determination, ROM_{Pitch} range of motion of pitch, *SVM* support vector machine, *SVR* support vector regression, *Trunk max acc_{AP}* maximum antero-posterior linear acceleration

[a]By definition, multivariate regression modelling did not feature distinct classes; rather, regression algorithms were utilised for problems featuring continuous and numeric outputs [94]

Table 6 Studies using motion-based measurements to assess movement quality

Study	Participants	Sensors and placements	Qty	Application	Measurement method	Gold/reference standard	Findings
Ahmadi et al. (2015) [59]	10 adults (9 healthy, 1 with low back pain; sex and age not specified)	Tri-axial accel + gyro (shank L&R, thigh L&R,	6	Joint angle measurement when jogging	Sensor orientations calculated using the Madgwick algorithm [48] with joint angles determined based on sensor orientations on proximal and distal segments	Joint angle measurements generated using normative data and applying a phase shift registration algorithm	The unregistered mean knee angle curve demonstrated significantly higher ($p=0.002$) magnitudes for 11%–17% and 87%–96% of the foot contact cycle and lower magnitudes ($p<0.001$), respectively, and for 61%–75% of the foot contact cycle. The unregistered mean hip angle curve had significantly lower ($p<0.010$) magnitudes for 14–22% and 74–86% of the foot contact cycle
Beange et al. (2019) [60]	10 healthy adults (6 males; males: 25.3 ± 2.2 years, females: 22.8 ± 2.2 years)	Tri-axial accel + gyro + mag (mid back, low back)	2	Measurement of lumbar spine flexion and extension	Sensor orientations calculated using a Kalman filter with flexion–extension angles obtained based on orientations of each sensor relative to one another	10-camera optical motion capture system (Vicon VantageV5, Vicon Motion Systems Ltd., Oxford, UK)	ICC when using both the FE angle time-series, and the SS time-series using the FE angle, lateral bend angle, and axial twist angle, to measure local dynamic stability ($0.807 \leq ICC_{FE} \leq 0.919$; $0.738 \leq ICC_{SS} \leq 0.868$) ICC for sagittal plane lumbopelvic coordination ($0.961 \leq ICC_{MARP} \leq 0.963$), ICC for sagittal plane lumbopelvic variability ($0.961 \leq ICC_{DP} \leq 0.963$)
Cortesi et al. (2019) [63]	14 healthy adults (14 males; 23.2 ± 2.8 years)[a]	Tri-axial accel + gyro + mag (chest, upper arm L&R, forearm L&R)	5[b]	Measurement of swimming stroke wrist trajectory	Wrist trajectories were determined using the Madgwick algorithm [48] coupled with the kinematic chain model [95]	7-camera optical motion capture system (BTS SMART-DX 7000, BTS Bioengineering, Milan, Italy)	RMSE between optical motion capture and sensor wrist trajectory = 7.7 cm 3D mean distance between optical motion capture wrist trajectory and sensor wrist trajectory = 13.0 ± 4.5 cm Correlation (r) between optical motion capture and the wearable sensors = 0.94

Table 6 (continued)

Study	Participants	Sensors and placements	Qty	Application	Measurement method	Gold/reference standard	Findings
Del Din et al. (2016) [64]	60 adults (30 healthy, 30 with PD; sex not specified; healthy: 66.6±7.7 years, PD: 66.9±9.4 years)	Tri-axial accel (low back)	1	Measurement of gait step length	Step length measured using the inverted pendulum model [89]	7.0-m long × 0.6-m wide instrumented walkway (Platinum model GaitRite, software version4.5, CIR systems, NJ, USA) and a synchronised webcam (Logitech, Webcam Pro 9000, CA, USA)	ICC between measurement accelerometer and instrumented walkway was 0.913 for healthy participants and 0.869 for those with PD
Lin et al. (2021) [71]	13 adults (10 healthy, 8 males; 3 with FS, sex not specified; 23.3±1.3 years)	Tri-axial accel+gyro+mag (upper arm L, forearm L, chest)	3	Measurement of shoulder ROM	Shoulder ROM calculated using a complementary filter to determine rotation angles	Optical motion capture system (Vicon Motion Systems Ltd. Oxford, UK)	Mean RMSE for all ROM tests ranged from 2.5° to 3.6°
Liu et al. (2020) [72]	6 (sex, age and medical status not provided)	Tri-axial accel+gyro+mag (upper arm L&R, forearm L&R, low back, mid back, thigh L&R, shank L&R, foot L&R)	12[c]	Joint angle measurement during a canoeing stroke	Sensor orientations calculated using the Madgwick [48] algorithm with joint angles determined based on sensor orientations on proximal and distal segments	12-camera optical motion capture system (OptiTrack, Natural Point Inc. USA)	Mean error for joint angles ranged from 1.20 to 3.72%
Meng et al. (2019) [74]	10 healthy adults (6 males; 26.5±6.2 years)	Tri-axial accel+gyro (shank, foot)	2	Measurement of ankle angle when walking	Orientation of each sensor is calculated using a complementary filter where the ankle joint angles are subsequently calculated	12-camera optical motion capture system (Vicon MX Giganet, Vicon Motion Systems Ltd. Oxford, UK)	RMSE of the sensors compared to the optical motion capture system was <3.5° for all measured walking speeds
Michaud et al. (2021) [75]	39 healthy adults (21 males; 39±21 years)	Tri-axial accel+gyro+mag (low back, mid back)	2	Measurement of lumbar flexion and pelvic tilt when performing hip hinge exercises	Sensor orientations calculated using the Madgwick [48] algorithm with lumbar flexion angles estimated based on sensor orientations	8-camera optical motion capture system (OptiTrack FLEX 3, Natural Point Inc. USA)	Lumbar flexion (2 IMUs): RMSE of 2.9° and 2.5° throughout movement and 2.7° and 3.0° at maximum angles for American kettlebell swing and deadlift, respectively. Pelvic tilt (1 IMU): RMSE of 2.4° and 1.9° at maximum angles for American kettlebell swing and deadlift, respectively

Table 6 (continued)

Study	Participants	Sensors and placements	Qty	Application	Measurement method	Gold/reference standard	Findings
Mitternacht et al. (2022) [76]	5 healthy adults (5 males; 25–61 years)	Tri-axial accel + gyro (shank)	1	Measurement of motor control during a battery of functional movement tests[d]	Kinematic variables obtained through integration of short sections of IMU time series data extracted from longer analysis sequences	8-camera optical motion capture system (Simi Reality Motion Systems GmbH, Unterschleissheim, Germany)	One-leg squat test for knee valgus: RMSE = 1.4° Drop jump squat test for knee valgus: RMSE = 2.2°, ΔX was 47% lower for the IMU (8.6 mm) than the camera system (16.2 mm)
Shepherd et al. (2017) [79]	10 healthy adults (0 males)	Tri-axial accel + gyro + mag (forearm)	1	Measurement of forearm angle during a netball shot	Madgwick [48] algorithm used to determine the pitch angle of the wearable device relative to the Earth's magnetic frame of reference from which the forearm angle was derived	10-camera optical motion capture system (Vicon VantageV16, Vicon Motion Systems Ltd. Oxford, UK)	The sensors were found to overestimate the forearm angle by 4.03 ± 1.58% on average across all shots compared to the optical motion capture system
Shuai et al. (2022) [80]	20 healthy adults (10 males; 25.19 ± 2.8 years)	Tri-axial accel + gyro + mag (low back, thigh L&R, shank L&R, foot L&R)	7	Measurement of hip, knee, and ankle angles during a battery of functional tests	Sensor orientations calculated using a Kalman filter with joint angles obtained based on orientations of each sensor relative to one another	8-camera optical motion capture system (OptiTrack Prime17W, Natural Point Inc. USA)	Ankle RMSE before zero offsetting: hip 4.7°–9.3°, knee = 3.6°–11.0°, ankle = 4.4°–13.1° Ankle RMSE after zero offsetting: hip = 3.0°–7.5°, knee = 2.1°–10.0°, ankle = 3.4°–12.8°
Tulipani et al. (2018) [83]	10 healthy adults (5 males)	Tri-axial accel + gyro (low back, pelvis, thigh L&R, shank L&R)[e]	6	Measurement of angular displacement of body segments during functional tasks	Angular velocity data from gyroscopes were integrated to obtain angular displacement using an undisclosed algorithm	9-camera optical motion capture system (Qualisys AB, Gothenburg, Sweden)	Overall average RMSE of the sensors compared to the optical motion capture system for the for angular displacement across all tasks and segments was 3.7° (RMSE range = 2.0°–7.1°)

Findings are reported to the degree of accuracy provided in the original study

3D three-dimensional, *accel* accelerometer, *DP* deviation phase, *FE* flexion–extension, *FS* frozen shoulder, *gyro* gyroscope, *ICC* intra-class correlation coefficient, *L* left, *mag* magnetometer, *MARP* mean absolute relative phase, *PD* Parkinson's disease, *R* right, *r* correlation, *RMSE* root-mean-square error, *ROM* range of motion, *SS* sum of squares, *ΔX* mean medial shift

[a]The phase of study that was applicable for the systematic review only featured five participants from the total sample

[b]While five sensors were worn in the study, only one device was worn when measuring the wrist trajectory

[c]The authors suggested that not all sensors were used in the measurement of joint angles in this study, stating that only sensors on the upper limbs were considered. However, it is not clear which of the sensors were actually used

[d]While the study featured a battery of 10 movement tests, only two movements met the inclusion criteria for this systematic review: one-leg squat and drop jump

[e]Only angular velocity data were presented in the paper, suggesting the use of the gyroscope exclusively. However, the accelerometer may have been used in the calibration and cannot be omitted without further information

[68], k-means clustering [65, 69], logistic regression [72], decision tree [72], XGBoost [72], random forests [77, 78], CNN [81, 82], and LSTM [81, 82]. Three studies employed multiple classifiers to determine the highest-performing methods for their respective applications [68, 72, 82]. Across the range of classifiers, the studies employed different approaches to feature selection (Table 5), although, with the exception of Spilz and Munz [81], the studies typically utilised raw sensor time-series data ($n = 15$). However, it was also common to derive kinematic data from the raw data ($n = 3$) [67, 73, 85] and capitalise on sensor fusion ($n = 6$) [66, 68, 72, 77, 78, 82] to expand the available features, while also extracting features from the raw data in the frequency domain ($n = 6$) [66, 68, 72, 77, 78, 82]. An assortment of validation methods was applied to the classifiers throughout all included studies; six studies used a real-world validation approach based on video footage [61, 62, 65, 67, 69, 70]. However, statistical methods were typically preferred, with seven studies using a k-fold cross-validation (KF-CV) [66, 68, 72, 81, 82, 84, 85], four a leave-one-subject-out cross-validation (LOSOCV) [68, 77, 78, 81], and two a leave-one-out cross-validation (LOOCV) [73, 77].

3.4 Methods and Reference Standards for the Measurement of Kinematic Characteristics to Assess Movement Quality

Sensor-fusion algorithms enable the calculation of sensor orientation, providing details of angular rotation about each of the three axes when considering movement in three-dimensional space. When wearable devices are positioned either side of a joint, or joints, the orientations relative to one another can enable joint angles to be estimated. A sensor-fusion approach was taken in 9 of the 12 measurement studies, including Cortesi et al. [63], where the kinematic chain model was used in conjunction with sensor fusion to generate both rotational and translational information [95]. Shepherd et al. [79] used sensor fusion in a simpler concept by calculating forearm angles using the orientation of a single device relative to the ground, comparable with Lin et al. [71], who utilised certain components of a shoulder range of motion (ROM) assessment. Similarly, Tulipani et al. [83], using gyroscope data exclusively, determined angular displacement for body segments rather than joint angles, a method also adopted by Mitternacht et al. [76], who measured tibial tilt angles for assessing knee valgus. However, Mitternacht et al. [76] also identified the linear medial shift of the knee joint by double-integrating the mediolateral acceleration obtained from the accelerometer. Del Din et al. [64] was the only study to exclusively measure translational movement without rotation or orientation, also using an integration-based method to obtain displacement and thereby calculate the length of a step.

To validate measurements of motion using novel instrumentation or methods, a predetermined reference standard is typically utilised for comparison, as is the case for all 12 of the measurement-based studies included in this review (see Table 6). Del Din et al. [64], for example, used an instrumented walkway to validate the measurement accelerometer, providing a strong intra-class correlation coefficient (ICC) for both healthy participants (ICC = 0.913) and for those with Parkinson's disease (ICC = 0.869). However, optical motion capture, generally considered to be the gold-standard for measuring movement [22], was utilised by 10 of the 12 measurement studies to validate the wearable sensor measurements (Table 6) [60, 63, 71, 72, 74–76, 79, 80, 83]. Six studies evaluated the accuracies of the angles calculated using the wearable sensor data in terms of root-mean-square error (RMSE) when compared to the optical motion capture systems [71, 74, 75, 76, 80, 83]. Considering three different walking speeds (0.5, 1.0 and 1.5 m·s^{-1}), Meng et al. [74] reported that the RMSE was always less than 3.5°. This is congruent with Tulipani et al. [83], where the overall average RMSE across a series of movements was 3.7°, and Lin et al. [71], who observed a maximum mean RMSE of 3.6° across the series of tests conducted. Moreover, Michaud et al. [75] found similar results when monitoring lumbar flexion and pelvic tilt, with RMSE ranging from 1.9° to 3.0° for the barbell deadlift and the American kettlebell swing. Mitternacht et al. [76] also reported comparable results, where the RMSE ranged from 1.4° to 2.2°. However, whilst the measurements were captured to determine knee instability during movements featuring knee flexion, namely the one-leg squat and drop jump, knee instability was not evident, resulting in minimal movement of the IMU. Consequently, the observed measurement errors were small [76]. Shuai et al. [80] reported a greater degree of error, with RMSE ranging from 2.1° to 13.1° across all measured joints and configurations. Cortesi et al. [63] evaluated the translational motion of a wrist-worn device against an optical motion capture system and reported a RMSE of 7.70 cm. Both Shepherd et al. [79] and Liu et al. [72] stated the mean error as percentages, indicating that the sensors overestimated the forearm angle by an average of 4.03% and calculating a maximum mean error of 3.72%, respectively. Beange et al. [60] reported on the reliability of wearable sensors when compared against an optical motion capture system for measuring spinal motion. Specifically, the authors reported on spinal flexion–extension angles ($0.807 \leq ICC_{FE} \leq 0.919$), and the sum of squares of flexion–extension, lateral bend, and axial twist angles to measure local dynamic stability ($0.738 \leq ICC_{SS} \leq 0.868$) [60]. Ahmadi et al. [59] utilised normative data and applied a phase-shift registration algorithm [96] to refine the joint angle measurements before conducting an

intra-study comparison against an unregistered approach. The phase shift, a pre-analysis curve transformation technique that uses a timescale shift based on the position of signal features [96], was used to align flexion–extension curves for foot contact cycles during jogging. Both the registered and unregistered curves were stated by Ahmadi et al. [59] to be representative of the mean joint angles obtained across the sample, where it was observed that the unregistered curve significantly underestimated the joint angle maxima (p = 0.002) and minima (p < 0.001) relative to the phase-shift registration algorithm. While significant, with the authors noting that even small differences could be indicative of injury, or increased risk of injury, they also acknowledged observed similarities between the joint angle curves generated throughout the duration of each movement.

4 Discussion

This systematic review sought to highlight any additional benefits that could be gained using multi-sensor devices, or multiple wearable devices, in place of unimodal sensors or a single device, respectively, when assessing movement quality. Accordingly, this review also investigated the additional variables, and indeed features, that could be obtained through multi-sensor devices for use in movement analyses. Further, the systematic review aimed to differentiate between the processing methods and applications of multi-sensor wearable devices in comparison to unimodal sensors when assessing movement quality. Evidence from the current review suggests that most movement quality assessments utilising wearable technology centre around expert-led or expert-based assessments [61, 62, 67–70, 73, 76–81, 83], often capitalising on pre-validated movement screening methods [61, 67, 69, 73, 76, 81]. Alternatively, movements are commonly assessed using proficient performers as the baseline [59, 60, 64, 66, 71, 72, 75, 84, 85]. Nevertheless, there are additional methodological considerations for the assessment of movement quality, and the authors of future reviews are recommended to further consider the specifics surrounding movement-quality assessment methods and the integration of technology. The reviewed studies revealed two overarching themes in the use of wearable technology to assess movement quality, with technology either used to classify movements or to directly measure motion for comparison against a baseline. To distinguish between the two methods, the results of these studies were reported independently, though Liu et al. [72] featured in both measurement and classification categories. It is pertinent to note that accuracies reported for each included study may not be

comparable if applied to other movements with alternative sensor placements.

All studies included in the review utilised an accelerometer, a gyroscope, or a magnetometer, or a combination thereof. Accelerometers were the most common of the three sensors, likely due to the versatility and historical application of these sensors when measuring movement [97]. Most often, the sensors were tri-axial and could therefore capture data in three dimensions, which is desirable given that typical human movements are not linearly constrained. Nonetheless, Ghasemzadeh and Jafari [65] used a tri-axial accelerometer coupled with a bi-axial gyroscope, although the two-dimensional angular velocity data obtained from the gyroscope were most likely sufficient for the intended application, as the focus was solely on rotation in the transverse plane.

Across all studies, the sampling frequency did not vary substantially when using a classification or measurement method. The comparable sampling frequencies are possibly related to the type of movements assessed, all of which were gross motor skills. Additionally, the range of sampling frequencies follows existing trends in other areas of the literature for similar applications [21]. Higher sampling frequencies may increase the clarity of motion measurements, particularly in explosive or finer movement patterns [40], although no included study considered fine motor skills and only five of the included studies could be considered to have used explosive movements: a baseball bat swing [65], overarm throw [67], table tennis forehand strikes [82], drop jump [76], and countermovement jump [80]. These studies featured sampling frequencies of 50 Hz, 128 Hz, 70 Hz, 200 Hz, and 100 Hz, respectively, which are relatively low in contrast to the frequencies recommended by Worsey et al. [40], and mostly at the lower end of the range of sampling frequencies observed in this review. Two of these studies [65, 67] successfully utilised a true classification method, suggesting that, for each classification method applied, the sampling frequencies were adequate. Tabrizi et al. [82] used a regression method, which was also successfully implemented using a sampling frequency below the average across this review. While Mitternacht et al. [76] and Shuai et al. [80] both aimed to measure motion during explosive movements, each study has limitations. Specifically, none of the participants in the study by Mitternacht et al. [76] exhibited noticeable knee instability during the drop jump, suggesting that observed tibial movement was minimal. Shuai et al. [80] used a height-restricted countermovement jump to prevent marker occlusion as well as timing controls to reduce the influence of movement speeds on measurements. Consequently, it was not possible based on the evidence provided by the studies included in this review to deduce whether an increased sampling frequency would have an inherent benefit

for measuring motion in studies assessing explosive movements, though speculatively, high sampling frequencies may be required to obtain accurate instantaneous measurements for high-speed movements due to the risk of aliasing error [98]. Indeed, it is not uncommon for average measurements to be utilised when measuring fast movements using wearables [99] due to the inability of low-sampling-frequency devices to capture sufficient data during rapid changes in the signal [98]. Yet, what we can determine from this review is that it is feasible to assess movement quality, including rapid human movements, using comparatively lower sampling frequencies when employing classification methods, for which high signal resolution may not be required. Of note, sensors with enhanced design features typically have higher unit costs [100], meaning most cost-effective, commercially available sensors likely avoid particularly high sampling frequencies.

The included studies featuring a classification approach highlighted the variety of factors that can influence the accuracy of classifiers. O'Reilly et al. [77, 78], for example, drew intra-study comparisons, where it was observed that, overall, a decrease in the quantity of sensors used tended to reduce classification accuracy. However, it was reported in both studies that the accuracy was not always markedly diminished by reducing the number of sensors utilised, depending on sensor location. Specifically, it was shown that single sensors worn on the shanks during bodyweight squats [78], or on the lower back or thighs for deadlifts [77] could provide comparable classification accuracies to a configuration with as many as five sensors. However, classification accuracy typically reduced considerably when using fewer sensors in sub-optimal positions for the respective movements, such as the shanks for the deadlift [77] and the lower back for bodyweight squats [78]. This is further evidenced when comparing the results of the two O'Reilly et al. [77, 78] studies, with the multi-class classifier employed for assessing squat technique [78] being noticeably more accurate than the equivalent classifier for the deadlift [77], despite the same sensor placements being used. It is speculated that this is due to the more proximal positions of the sensors to the main area of the observed deviations (i.e., the lower limbs) during the squat, whereas the deviations identified for the deadlift largely overlooked lower limb movements and positions. However, it may be feasible to obtain good classification accuracies with multi-class classifiers using fewer sensors. As indicated by Masci et al. [73], as few as one sensor may indeed be adequate when assessing global movements provided that broader criteria are employed, rather than aiming to highlight a specific movement discrepancy. However, further research is required to further explore this hypothesis. Nonetheless, it is essential to recognise that when using fewer than the maximal available sensors, the best-performing device positions were dependent

on the movements assessed by O'Reilly et al. [77, 78]. This is a limitation for real-world applications, as to use the optimal sensor position for each movement would require an adjustment to the anatomical positioning of a sensor, or sensors, to maximise accuracy when conducting a sequential assessment with multiple movements, a potentially time-consuming step. However, it may be plausible to consider a practical compromise by using a single-sensor placement that ensures adequate accuracy when assessing multiple movements, without rigidly adhering to the optimal position for each specific movement. This approach can help alleviate the requirement for time-consuming adjustments to individual sensor positions, making it more feasible and efficient for real-world applications. O'Reilly et al. [77, 78] also conducted intra-study comparisons between binary classifiers, where movements were categorised as either proficient or not proficient, and a multi-class classifier featuring five classes, where the specific deviations from the accepted movement standard were highlighted. Congruent with Kianifar et al. [68], the multi-class classifiers were less accurate than the binary classifiers due to the need to distinguish specific movement errors from the data [77, 78]. However, multi-class classifiers have the potential to be more informative, given that specific movement discrepancies can be detected [77, 78]. Future research to improve the accuracy of multi-class classifiers is therefore warranted. Notably, O'Reilly et al. [77, 78] and Kianifar et al. [68] considered movements that were lower-limb dominant, with the sensor configurations reflecting this. A key consideration of this review is that the accuracies reported in each study may not be comparable if applied to other movements with alternative sensor placements. Moreover, classifiers based on natural movement deviations appear to underperform in comparison to those where movement errors were induced [77]. This highlights a key challenge that may arise when performing movement quality assessments using wearable devices in real-world applications.

While O'Reilly et al. [77, 78] and Kianifar et al. [68] considered the relationship between a reduction in the quantity of sensors and the accuracy of classifiers, each placement location featured a device consisting of more than one type of sensor. This is because each study utilised sensor fusion to provide additional metrics, such as limb orientations and joint angles, which were fed into the respective classification algorithms [68, 77, 78], as was the case with Ghobadi and Esfahani [66], Liu et al. [72], and Tabrizi et al. [82]. Unfortunately, based on the reviewed literature, it is not possible to make comparisons between the use of multi-sensor devices and unimodal sensors in equivalent anatomical positions, nor to consider the use of multiple types of unimodal sensors applied outside of a single device when assessing movement quality. Indeed, the methodological heterogeneity among the included studies prevents conclusions being drawn regarding

the potential benefits of using multi-sensor devices over uni-modal alternatives. It is speculated that unimodal sensors would not perform as well as multi-sensor devices for identifying movement discrepancies using classifiers, not least because it would only be possible to obtain the orientation data, and subsequently utilise the orientation signal features if sensor combinations exist within one device. In such instances where additional signal features were obtained, classifiers were often very accurate [66, 72, 82], and possibly more so than those with fewer features. However, lower accuracies were also detected using the same signal features under less optimal configurations [68, 77, 78].

Accelerometers were used in isolation in four of the classification studies [62, 69, 70, 84]. Of interest, accelerometers were the only type of sensor used in isolation and most commonly in studies that required classifiers with less complexity. Specifically, Lee et al. [70] and Caporaso and Grazioso [62] used algorithmically simple classifiers based on temporal features of the acceleration signal to detect loss of ground contact during race walking, while Lander et al. [69] and Xu et al. [84] also considered the features of acceleration signals but assessed their correlation with pre-existing signals that corresponded with key assessment criteria. The absence of additional sensors appears to restrict what may be achieved, and while this is not always a barrier when implementing simple classifiers, other studies using additional sensory data have demonstrated greater detail, specificity, and applicability to a wider selection of movements. Indeed, sensor fusion, utilised in six of the classification studies [48, 66, 68, 72, 77, 78, 82] to obtain orientation information as an additional input to the classifier, was reported to be associated with good accuracies.

It is presently unclear as to how influential the specific assessment criteria and movements are on the accuracy of classifiers, meaning it is difficult to definitively state whether additional sensor types within a device can improve classifier accuracy. It was theorised that multi-sensor devices would increase classifier accuracy due to the increase in measurable outputs, though this would only be possible to demonstrate through direct comparisons. However, it is pertinent to note that key a priori decisions, such as the quantity and configuration of sensors, may influence the apparent classifier accuracy and its interaction with protocol-related factors, for example, movements assessed and sensor placement. This is suggested by the comparable binary classifier accuracies presented by Caporaso and Grazioso [62] and Lee et al. [70], who both used a single accelerometer on the lower back, and Kianifar et al. [68], in whose study a single multi-sensor device was worn on the shank. The classifier accuracies in all three studies were determined by comparing against video footage assessed and labelled by an expert [62, 68, 70]. Both Kianifar et al. [68] and Caporaso and Grazioso [62] also employed three-level classifiers, which

also indicated comparable accuracies. Notably, Kianifar et al. [68] applied classification methods to a unilateral squat, whereas Caporaso and Grazioso [62] and Lee et al. [70] used classifiers to highlight race-walking infringements. It is feasible that the increased instability that arises with unilateral movements, such as the unilateral squat used by Kianifar et al. [68], could have an influence on the assessment accuracy, offsetting the possible benefits of additional sensors. To reinforce the capability of single-sensor units, Xu et al. [84] presented near perfect classification accuracy when measuring a cycling pedal motion, despite using a single accelerometer. As the motion assessed was constrained to a fixed path, less variation in the movement was possible, thereby potentially simplifying the detection of key signal features.

Traditional feature-based classifiers were a common choice [61, 62, 65, 67, 70, 73], although no indication was given as to why this approach was selected over other classifiers. It is possible that the simplicity of the assessment may be one factor [62, 70] while the benefits of using temporal parameters may be preferred when considering movement sequencing [65]. It does appear, however, that highlighting specific movement discrepancies is lacking in these studies due to a reliance on broad assessment criteria [61, 67, 73]. A further limitation of traditional feature-based classifiers is a reliance on real-world validation through manual assessment, a source of subjectivity and human error [61, 62, 65, 67, 70]. Taken together, it is speculated that supervised machine-learning algorithms are preferable where possible, as they often use the same common signal features for the analysis but can negate much of the human error by removing subjectivity through the addition of algorithm training [66, 68, 72, 77, 78, 81, 82, 84, 85].

While an array of supervised learning methods was considered, SVM was the most widely used. This is unsurprising given that SVM consistently outperforms other classifiers in the reviewed literature, as demonstrated by intra-study comparisons with other methods [68, 72]. Notably, in the only study that utilised multivariate regression [82], SVR, a regression model based on SVM, was shown to have comparable results for quantifying movement quality when evaluated against deep-learning alternatives, namely LSTM and CNN [82]. Indeed, while LSTM marginally outperformed both SVR and CNN models, albeit all models were fit for purpose, SVR may be more optimal than LSTM when using smaller datasets [82]; traditional machine learning algorithms often generate comparable outcomes to deep learning methods under such circumstances [101]. Accordingly, Kianifar et al. [68] specifically highlighted the utility for smaller datasets and high-dimensional data that is characteristic of most of the included classification studies. It is important to recognise, however, that while SVM appears to be preferable in the assessment of movement quality, no

machine learning algorithm is uniformly superior under all conditions [102, 103]. Indeed, the wider body of literature illustrates both the benefits and limitations that exist for the numerous machine learning options [102–104]. Hence, it is essential that researchers consider the array of available options when determining the most suitable approach for their intended application.

When employing supervised and unsupervised machine-learning methods, statistical validation methods were generally used in lieu of manual assessments, with KF-CV and LOSOCV being the two primary options. Kianifar et al. [68] utilised and compared both validation methods, with KF-CV providing more accurate results. However, other studies employed LOSOCV with near-perfect accuracy [78], the suitability of which was specifically recommended by O'Reilly et al. [77] for universal classifiers. In contrast, O'Reilly et al. [77] employed LOOCV, as utilised in Masci et. al. [73], to validate a personalised classifier, where the classifier incorporated user-specific details. While this did improve accuracy considerably, the limitation of personal classifiers is the need to adapt to each user, introducing greater time demands and a less user-friendly experience. The most applicable validation method may therefore be dependent on a range of factors, such as the application, sensor positions, and data type. A notable example of this is in the study by Spilz and Munz [81], who implemented a CNN-LSTM layered neural network with two different validation methods, LOSOCV and KF-CV, at different stages of the network architecture.

Whilst unimodal sensor data has been used to assess movement quality [64], sensor-fusion algorithms are particularly prevalent in measurement-based studies. This is likely due to the issues associated with obtaining kinematic data from processing unimodal data, such as integration drift and gimbal lock. Indeed, the application of sensor fusion helps overcome such issues, whereby each sensor compensates the limitations of other sensors [32]. However, sensor fusion is not a complete solution for assessing movement quality when implemented with IMU data as it does not enable the derivation of linear kinematic variables, such as linear displacement and velocity. The measurement of linear motion, therefore, remains reliant on the manipulation of accelerometer data exclusively, which are typically erroneous [1, 32] unless combined with measurements from another measurement system, such as radio-based systems and cameras [32]. Nonetheless, several angular kinematic parameters can be gleaned from the sensor fusion of IMU data alone to highlight specific movement discrepancies [59, 60, 72, 74, 79]. Interestingly, Del Din et al. [64] performed a double integration of accelerometer data to obtain step length, a metric based on linear displacement, reporting that pre-processing using a high-pass Butterworth filter largely addressed the issue of integration drift, though it is pertinent

to note that other errors likely persisted [46]. Interestingly, the ICC was excellent between measurement methods, although Del Din et al. [64] acknowledged the limitations of using an instrumented walkway for comparison. Therefore, the findings are likely to be less reliable than studies implementing an optical motion capture system, which is the case for the study by Ahmadi et al. [59] too, where the instrumentation is validated against itself, albeit using optimised configurations. Mitternacht et al. [76] also performed a double integration on the acceleration data captured by the accelerometers to calculate linear position, reducing the influence of integration drift by dividing longer time series data into shorter 0.1 s segments. However, the medial shift estimate was too small to be considered reliable and was reported to a degree of precision that is likely unattainable based on the findings of other research [105, 106]. Unsurprisingly, therefore, the mean medial shift calculated using the IMU was found to be 47% lower than that obtained by the optical motion capture system, a relatively large amount of error. Similarly, Mitternacht et al. [76] and Tulipani et al. [83] integrated gyroscope data to calculate the angular displacement of limb segments. While the results were mostly accurate and reliable in comparison to the gold-standard optical motion capture, it must be reiterated that the degree of tibial tilt observed during the movements in the study by Mitternacht et al. was almost negligible, while Tulipani et al. [83] applied an undisclosed algorithm to, at least in part, overcome the gyroscope deterministic bias offset, rather than simply integrating. Without disclosure of all processing methods, it is not possible to identify what other methods may have influenced the results, thereby limiting interstudy comparisons. It is also important to note that only planar motion was considered, restricting movement to two, rather than three, dimensions, which could be influential on the accuracies reported.

Within the reviewed measurement studies, and indeed all studies included in the review, the Madgwick algorithm [48] was the most frequently used method of sensor fusion. Relative to Kalman and complementary filters, the Madgwick algorithm is still novel. While the reporting of accuracies with each method and the selection of movements is largely heterogeneous, there were no obvious advantages of the Madgwick algorithm identified for achieving accurate measurements in comparison with Kalman and complementary filters. However, both Ahmadi et al. [59] and Shepherd et al. [79] justified their selection of the Madgwick algorithm based on the low-computational demands of the algorithm. Shepherd et al. [79] also identified the suitability of the algorithm when aiming to utilise lower sampling rates and reduce power consumption [59, 79], ideal qualities for incorporation in commercially available wearable devices. While Beange et al. [60] and Shuai et al. [80] did not provide any reasons behind the use of the Kalman filter, Meng et al.

[74] highlighted the capability of the complementary filter approach to sensor fusion to overcome drift. However, both the Madgwick algorithm and Kalman filters can also be used to overcome drift, so it is difficult to attribute the selection of a complementary filter on this basis alone. Other possible reasons for the selection of a complementary filter could be the absence of a magnetometer, or simply the ease of implementation [107]. It is also plausible that an undisclosed design feature was utilised as part of the complementary filter to optimise performance for their specific application, given that the authors allude to the use of complementary filters in similar studies also assessing gait [74].

An observed trend is the use of sensor fusion to estimate joint angles, achieved by placing devices on both proximal and distal, or inferior and superior, segments and determining the orientation of each device relative to another [59, 60, 71, 72, 74, 75, 80]. This method allows for the estimation of ROM, a particularly useful metric in both sports and clinical settings. Moreover, orientation data can also be applied relative to fixed coordinate systems. This approach was utilised by Shepherd et al. [79] to assess the forearm angle relative to the ground during a netball shot, and Cortesi et al. [63] to estimate wrist orientation throughout a swimming stroke as part of a more complete motion measurement. Both Shepherd et al. [79] and Cortesi et al. [63] utilise a single device for the measurement of specific components of an activity, which, whilst it may be adequate depending on the application, more complete movement assessments will typically require additional sensors. Indeed, even two sensors, the minimum requirement for estimating joint angles using sensor fusion, has been shown to be insufficient for systemic measurements [60]. Therefore, when using wearable devices, each unit is only appropriate for localised measurements, such as single-limb segments or single joints, thereby necessitating additional sensors to consider broader criteria.

5 Conclusion

In conclusion, this systematic review has highlighted some of the key differences between the applications and processing methods associated with the use of unimodal and multi-sensor wearable devices to assess movement quality. Further, the use of multiple devices increases the feasibility of effectively assessing holistic movements, while multi-sensor devices offer the ability to obtain more output metrics. Actions should be taken to further improve measurement accuracy and multi-class classification accuracy, and to translate the systems into affordable, accessible, real-world solutions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40279-023-01905-1.

Declarations

Conflict of interest TAS received funding, which provided this author with a PhD stipend, from a project sponsor, Polar Electro Oy. The authors declare that they have no further known competing financial interests or personal relationships that could have influenced the content of this article.

Availability of data and material All data generated or analysed during this study are included in this published article and its supplementary information files.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Code availability Not applicable.

Author contributions TAS, MAM and KAM made substantial contributions to the conception and design of the work. TAS and AWHR made substantial contributions to the acquisition of data for the work. All authors made substantial contributions to the analysis and interpretation of data for the work. TAS drafted the initial manuscript and contributed to the implementation of revisions to optimise the delivery of important intellectual content. MAM, KAM and AWHR also contributed to the content of the manuscript, revising it critically for important intellectual content. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work.

References

1. Camomilla V, Bergamini E, Fantozzi S, Vannozzi G. Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review. Sensors [Internet]. 2018 [cited 2022 Sep 15];18:873. https://www.mdpi.com/1424-8220/18/3/873.

2. Clark CCT, Barnes CM, Holton M, Summers HD, Stratton G. A kinematic analysis of fundamental movement skills. Sport Science Review [Internet]. 2016;25:261–75. http://archive.sciendo.com/SSR/ssr.2016.25.issue-3-4/ssr-2016-0014/ssr-2016-0014.pdf.

3. Clark CCT, Barnes CM, Holton M, Summers HD, Stratton G. Profiling movement quality and gait characteristics according to body-mass index in children (9–11 y). Hum Mov Sci [Internet]. 2016;49:291–300. https://linkinghub.elsevier.com/retrieve/pii/S0167945716301075.

4. Andrews JR, Fleisig GS. Preventing throwing injuries. Journal of Orthopaedic & Sports Physical Therapy [Internet]. 1998;27:187–8. https://doi.org/10.2519/jospt.1998.27.3.187.

5. Lyman S, Fleisig GS, Andrews JR, David Osinski E. Effect of pitch type, pitch count, and pitching mechanics on risk of elbow and shoulder pain in youth baseball pitchers. 2002.

6. Young RW. The ontogeny of throwing and striking. human_ontogenetics [Internet]. 2009;3:19–31. https://doi.org/10.1002/huon.200800013.

7. Daoud AI, Geissler GJ, Wang F, Saretsky J, Daoud YA, Lieberman DE. Foot strike and injury rates in endurance runners: A retrospective study. Med Sci Sports Exerc [Internet]. 2012 [cited 2023 Apr 6];44:1325–34. https://journals.lww.com/acsm-msse/Fulltext/2012/07000/Foot_Strike_and_Injury_Rates_in_Endurance_Runners_.17.aspx.

8. Dudagoitia E, García-de-Alcaraz A, Andersen LL. Safety of powerlifting: a literature review. Sci Sports [Internet]. 2021 [cited 2023 Apr 6];36:e59–68. https://linkinghub.elsevier.com/retrieve/pii/S0765159720302069.

9. Myer GD, Quatman CE, Khoury J, Wall EJ, Hewett TE. Youth Versus Adult "Weightlifting" Injuries Presenting to United States Emergency Rooms: Accidental Versus Nonaccidental Injury Mechanisms. Journal of strength and conditioning research/National Strength & Conditioning Association [Internet]. 2009 [cited 2023 Apr 6];23:2054. Available from: /pmc/articles/PMC4034275/.

10. Whitehead M, editor. Physical literacy: Throughout the lifecourse [Internet]. 1st ed. Physical Literacy: Throughout the Lifecourse. London: Routledge; 2010. https://www.taylorfrancis.com/books/9781134010684.

11. Zhang X, Shan G, Wang Y, Wan B, Li H. Wearables, biomechanical feedback, and human motor-skills' learning & optimization. Applied Sciences [Internet]. 2019;9:226. http://www.mdpi.com/2076-3417/9/2/226.

12. Kolman NS, Kramer T, Elferink-Gemser MT, Huijgen BCH, Visscher C. Technical and tactical skills related to performance levels in tennis: A systematic review. J Sports Sci [Internet]. 2019 [cited 2023 Apr 6];37:108–21. https://www.tandfonline.com/action/journalInformation?journalCode=rjsp20.

13. Basevitch I, Tenenbaum G, Land WM, Ward P. Visual and skill effects on soccer passing performance, kinematics, and outcome estimations. Front Psychol [Internet]. 2015 [cited 2023 Apr 6];6. Available from: /pmc/articles/PMC4345767/.

14. Bardid F, Vannozzi G, Logan SW, Hardy LL, Barnett LM. A hitchhiker's guide to assessing young people's motor competence: Deciding what method to use. J Sci Med Sport. Elsevier Ltd; 2019. p. 311–8.

15. Venek V, Kranzinger S, Schwameder H, Stöggl T. Human movement quality assessment using sensor technologies in recreational and professional sports: a scoping review. Sensors. 2022;22.

16. König N, Taylor WR, Baumann CR, Wenderoth N, Singh NB. Revealing the quality of movement: a meta-analysis review to quantify the thresholds to pathological variability during standing and walking. Neurosci Biobehav Rev. 2016;68:111–9.

17. Warner MB, Wilson DA, Herrington L, Dixon S, Power C, Jones R, et al. A systematic review of the discriminating biomechanical parameters during the single leg squat. Phys Ther Sport. 2019;36:78–91.

18. Houwen S, Hartman E, Jonker L, Visscher C. Reliability and validity of the TGMD-2 in primary-school-age children with visual impairments. Adapted Physical Activity Quarterly [Internet]. 2010;27:143–59. https://journals.humankinetics.com/view/journals/apaq/27/2/article-p143.xml.

19. Valentini NC, Rudisill ME, Bandeira PFR, Hastie PA. The development of a short form of the Test of Gross Motor Development-2 in Brazilian children: validity and reliability. Child Care Health Dev [Internet]. 2018;44:759–65. https://doi.org/10.1111/cch.12598.

20. O'Reilly M, Caulfield B, Ward T, Johnston W, Doherty C. Wearable inertial sensor systems for lower limb exercise detection and evaluation: a systematic review. Sports Medicine [Internet]. 2018;48:1221–46. https://doi.org/10.1007/s40279-018-0878-4.

21. Adesida Y, Papi E, McGregor AH. Exploring the role of wearable technology in sport kinematics and kinetics: A systematic review. Sensors [Internet]. 2019;19:1597. https://www.mdpi.com/1424-8220/19/7/1597.

22. van der Kruk E, Reijne MM. Accuracy of human motion capture systems for sport applications; state-of-the-art review. Eur J Sport Sci [Internet]. 2018;18:806–19. https://doi.org/10.1080/17461391.2018.1463397.

23. Atack AC, Trewartha G, Bezodis NE. A joint kinetic analysis of rugby place kicking technique to understand why kickers achieve different performance outcomes. J Biomech [Internet]. 2019;87:114–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S002192901930154X.

24. Simonsen D, Popovic MB, Spaich EG, Andersen OK. Design and test of a Microsoft Kinect-based system for delivering adaptive visual feedback to stroke patients during training of upper limb movement. Med Biol Eng Comput [Internet]. 2017;55:1927–35. https://doi.org/10.1007/s11517-017-1640-z.

25. Schlagenhauf F, Sreeram S, Singhose W. Comparison of Kinect and Vicon motion capture of upper-body joint angle tracking. In: 2018 IEEE 14th international conference on control and automation (ICCA). 2018. p. 674–9.

26. Lourenço J, Martins L, Almeida R, Quaresma C, Vieira P. Low cost inertial measurement unit for motion capture in biomedical applications. IFIP Adv Inf Commun Technol [Internet]. Springer New York LLC; 2016 [cited 2020 Nov 30]. p. 151–8. https://doi.org/10.1007/978-3-319-31165-4_16.

27. Robert-Lachaine X, Mecheri H, Muller A, Larue C, Plamondon A. Validation of a low-cost inertial motion capture system for whole-body motion analysis. J Biomech [Internet]. 2020;99:109520. https://linkinghub.elsevier.com/retrieve/pii/S0021929019307742.

28. Van Hooren B, Goudsmit J, Restrepo J, Vos S. Real-time feedback by wearables in running: Current approaches, challenges and suggestions for improvements. J Sports Sci [Internet]. 2020;38:214–30. https://doi.org/10.1080/02640414.2019.1690960.

29. Clark CCT, Barnes CM, Duncan MJ, Summers HD, Stratton G. Physical activity, motor competence and movement and gait quality: A principal component analysis. Hum Mov Sci [Internet]. 2019;68:102523. https://linkinghub.elsevier.com/retrieve/pii/S0167945719304634.

30. Twomey N, Diethe T, Fafoutis X, Elsts A, McConville R, Flach P, et al. A comprehensive study of activity recognition using accelerometers. Informatics [Internet]. 2018;5:27. http://www.mdpi.com/2227-9709/5/2/27.

31. Allseits E, Kim K, Bennett C, Gailey R, Gaunaurd I, Agrawal V. A novel method for estimating knee angle using two leg-mounted

gyroscopes for continuous monitoring with mobile health devices. Sensors [Internet]. 2018;18:2759. http://www.mdpi.com/1424-8220/18/9/2759.

32. Kok M, Hol JD, Schön TB. Using inertial sensors for position and orientation estimation. Foundations and Trends in Signal Processing [Internet]. 2017;11:1–153. http://www.nowpublishers.com/article/Details/SIG-094.

33. Morrow MMB, Lowndes B, Fortune E, Kaufman KR, Hallbeck MS. Validation of inertial measurement units for upper body kinematics. J Appl Biomech [Internet]. 2017;33:227–32. https://journals.humankinetics.com/view/journals/jab/33/3/article-p227.xml.

34. Picerno P, Camomilla V, Capranica L. Countermovement jump performance assessment using a wearable 3D inertial measurement unit. J Sports Sci [Internet]. 2011;29:139–46. https://doi.org/10.1080/02640414.2010.523089.

35. Sgrò F, Mango P, Pignato S, Schembri R, Licari D, Lipoma M. Assessing standing long jump developmental levels using an inertial measurement unit. Percept Mot Skills [Internet]. 2017;124:21–38. https://doi.org/10.1177/0031512516682649.

36. Šlajpah S, Kamnik R, Munih M. Kinematics based sensory fusion for wearable motion assessment in human walking. Comput Methods Programs Biomed [Internet]. 2014;116:131–44. https://linkinghub.elsevier.com/retrieve/pii/S0169260713003842.

37. Ulrich DA. Test of gross motor development. 3rd ed. Austin, TX: Pro-ED; 2016.

38. Cook G, Burton L, Hoogenboom BJ, Voight M. Functional movement screening: the use of fundamental movements as an assessment of function—part 2. Int J Sports Phys Ther [Internet]. 2014 [cited 2023 Mar 27];9:549. Available from: /PMC4127517/.

39. Cook G, Burton L, Hoogenboom BJ, Voight M. Functional movement screening: the use of fundamental movements as an assessment of function—part 1. Int J Sports Phys Ther [Internet]. 2014 [cited 2023 Mar 27];9:396. Available from: /pmc/articles/PMC4060319/.

40. Worsey M, Espinosa H, Shepherd J, Thiel D. Inertial sensors for performance analysis in combat sports: a systematic review. Sports [Internet]. 2019 [cited 2022 Jul 6];7:28. http://www.mdpi.com/2075-4663/7/1/28.

41. Cust EE, Sweeting AJ, Ball K, Robertson S. Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. J Sports Sci [Internet]. 2019 [cited 2022 Jul 6];37:568–600. https://doi.org/10.1080/02640414.2018.1521769.

42. Saboor A, Kask T, Kuusik A, Alam MM, Le Moullec Y, Niazi IK, et al. Latest research trends in gait analysis using wearable sensors and machine learning: a systematic review. IEEE Access [Internet]. 2020 [cited 2022 Jul 6];8:167830–64. https://ieeexplore.ieee.org/document/9187883/.

43. Liu Q, Williamson J, Li K, Mohrman W, Lv Q, Dick RP, et al. Gazelle: Energy-efficient wearable analysis for running. IEEE Trans Mob Comput [Internet]. 2017 [cited 2022 Jul 8];16:2531–44. http://ieeexplore.ieee.org/document/7726063/.

44. Khan A, Hammerla N, Mellor S, Plötz T. Optimising sampling rates for accelerometer-based human activity recognition. Pattern Recognit Lett [Internet]. 2016 [cited 2022 Jul 8];73:33–40. https://linkinghub.elsevier.com/retrieve/pii/S0167865516000040.

45. Qiu S, Zhao H, Jiang N, Wang Z, Liu L, An Y, et al. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: state-of-the-art and research challenges. Information Fusion [Internet]. 2022 [cited 2022 Jul 20];80:241–65. https://linkinghub.elsevier.com/retrieve/pii/S1566253521002311.

46. Han S. Measuring displacement signal with an accelerometer. J Mech Sci Technol [Internet]. 2010;24:1329–35. https://doi.org/10.1007/s12206-010-0336-1.

47. Kok M, Hol JD, Schön TB. An optimization-based approach to human body motion capture using inertial sensors. IFAC Proceedings Volumes [Internet]. 2014;47:79–85. https://linkinghub.elsevier.com/retrieve/pii/S147466701641596X.

48. Madgwick SOH, Harrison AJL, Vaidyanathan R. Estimation of IMU and MARG orientation using a gradient descent algorithm. In: 2011 IEEE international conference on rehabilitation robotics [Internet]. IEEE; 2011. p. 1–7. http://ieeexplore.ieee.org/document/5975346/.

49. Papi E, Koh WS, McGregor AH. Wearable technology for spine movement assessment: a systematic review. J Biomech [Internet]. 2017;64:186–97. https://linkinghub.elsevier.com/retrieve/pii/S0021929017305109.

50. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med [Internet]. 2009 [cited 2020 Nov 5];6:e1000097. https://doi.org/10.1371/journal.pmed.1000097.

51. Boland A, Cherry MG, Dickson R. Doing a systematic review: a student's guide [Internet]. SAGE Publications; 2013. https://books.google.co.uk/books?id=AO-GAwAAQBAJ.

52. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev [Internet]. 2016;5:210. https://doi.org/10.1186/s13643-016-0384-4.

53. McHugh ML. Interrater reliability: The kappa statistic. Biochem Med (Zagreb) [Internet]. 2012;22:276–82. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/?report=printable.

54. Collins H, Booth JN, Duncan A, Fawkner S, Niven A. The effect of resistance training interventions on 'the self' in youth: a systematic review and meta-analysis. Sports Med Open [Internet]. 2019 [cited 2022 Sep 14];5:29. https://doi.org/10.1186/s40798-019-0205-0.

55. McDonagh M, Peterson K, Raina P, Chang S, Shekelle P. Avoiding Bias in Selecting Studies. Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]. Rockville, MD: Agency for Healthcare Research and Quality (US); 2013 [cited 2023 Mar 13]. https://www.ncbi.nlm.nih.gov/books/NBK126701/.

56. Maher C, Calia C. The effect of illiteracy on performance in screening tools for dementia: a meta-analysis. J Clin Exp Neuropsychol [Internet]. 2021 [cited 2022 Sep 14];43:945–66. https://doi.org/10.1080/13803395.2022.2040433.

57. Pai ALH, McGrady M. Systematic review and meta-analysis of psychological interventions to promote treatment adherence in children, adolescents, and young adults with chronic illness. J Pediatr Psychol [Internet]. 2014;39:918–31. https://doi.org/10.1093/jpepsy/jsu038.

58. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health (1978) [Internet]. 1998;52:377–84. https://doi.org/10.1136/jech.52.6.377.

59. Ahmadi A, Mitchell E, Richter C, Destelle F, Gowing M, O'Connor NE, et al. Toward automatic activity classification and movement assessment during a sports training session. IEEE Internet Things J [Internet]. 2015;2:23–32. http://ieeexplore.ieee.org/document/6981909/.

60. Beange KHE, Chan ADC, Beaudette SM, Graham RB. Concurrent validity of a wearable IMU for objective assessments of functional movement quality and control of the lumbar spine. J Biomech [Internet]. 2019;97:109356. https://linkinghub.elsevier.com/retrieve/pii/S0021929019305809.

61. Bisi MC, Pacini Panebianco G, Polman R, Stagni R. Objective assessment of movement competence in children using wearable sensors: An instrumented version of the TGMD-2 locomotor subtest. Gait Posture [Internet]. 2017;56:42–8. https://linkinghub.elsevier.com/retrieve/pii/S0966636217301571.

62. Caporaso T, Grazioso S. IART: Inertial Assistant Referee and Trainer for race walking. Sensors [Internet]. 2020;20:783. https://www.mdpi.com/1424-8220/20/3/783.

63. Cortesi M, Giovanardi A, Gatta G, Mangia AL, Bartolomei S, Fantozzi S. Sensors in swimming: detection of stroke phases through 3D wrist trajectory. J Sports Sci Med [Internet]. 2019;18:438–47. http://www.jssm.org.

64. Del Din S, Godfrey A, Rochester L. Validation of an accelerometer to quantify a comprehensive battery of gait characteristics in healthy older adults and Parkinson's disease: Toward clinical and at home use. IEEE J Biomed Health Inform [Internet]. 2016;20:838–47. http://ieeexplore.ieee.org/document/7078919/.

65. Ghasemzadeh H, Jafari R. Coordination analysis of human movements with body sensor networks: A signal processing model to evaluate baseball swings. IEEE Sens J [Internet]. 2011;11:603–10. http://ieeexplore.ieee.org/document/5482088/.

66. Ghobadi M, Esfahani ET. A robust automatic gait monitoring approach using a single IMU for home-based applications. J Mech Med Biol [Internet]. 2017;17:1750077. https://doi.org/10.1142/S0219519417500774.

67. Grimpampi E, Masci I, Pesce C, Vannozzi G. Quantitative assessment of developmental levels in overarm throwing using wearable inertial sensing technology. J Sports Sci [Internet]. 2016;34:1759–65. https://doi.org/10.1080/02640414.2015.1137341.

68. Kianifar R, Lee A, Raina S, Kulic D. Automated assessment of dynamic knee valgus and risk of knee injury during the single leg squat. IEEE J Transl Eng Health Med [Internet]. 2017;5:1–13. https://ieeexplore.ieee.org/document/8089739/.

69. Lander N, Nahavandi D, Mohamed S, Essiet I, Barnett LM. Bringing objectivity to motor skill assessment in children. J Sports Sci [Internet]. 2020;38:1539–49. https://doi.org/10.1080/02640414.2020.1747743.

70. Lee J, Mellifont R, Burkett B, James D. Detection of illegal race walking: a tool to assist coaching and judging. Sensors [Internet]. 2013;13:16065–74. http://www.mdpi.com/1424-8220/13/12/16065.

71. Lin Y-C, Tsai Y-J, Hsu Y-L, Yen M-H, Wang J-S. Assessment of shoulder range of motion using a wearable inertial sensor network. IEEE Sens J [Internet]. 2021;21:15330–41. https://ieeexplore.ieee.org/document/9405640/.

72. Liu L, Qiu S, Wang Z, Li J, Wang J. Canoeing motion tracking and analysis via multi-sensors fusion. Sensors [Internet]. 2020;20:2110. https://www.mdpi.com/1424-8220/20/7/2110.

73. Masci I, Vannozzi G, Bergamini E, Pesce C, Getchell N, Cappozzo A. Assessing locomotor skills development in childhood using wearable inertial sensor devices: the running paradigm. Gait Posture [Internet]. 2013;37:570–4. https://linkinghub.elsevier.com/retrieve/pii/S0966636212003633.

74. Meng L, Martinez-Hernandez U, Childs C, Dehghani-Sanij AA, Buis A. A practical gait feedback method based on wearable inertial sensors for a drop foot assistance device. IEEE Sens J [Internet]. 2019;19:12235–43. https://ieeexplore.ieee.org/document/8822472/.

75. Michaud F, Pérez Soto M, Lugrís U, Cuadrado J. Lower back injury prevention and sensitization of hip hinge with neutral spine using wearable sensors during lifting exercises. Sensors [Internet]. 2021;21:5487. https://www.mdpi.com/1424-8220/21/16/5487.

76. Mitternacht J, Hermann A, Carqueville P. Acquisition of lower-limb motion characteristics with a single inertial measurement unit—validation for use in physiotherapy. Diagnostics. 2022;12.

77. O'Reilly M, Whelan D, Ward T, Delahunt E, Caulfield B. Classification of deadlift biomechanics with wearable inertial measurement units. J Biomech [Internet]. 2017;58:155–61. https://linkinghub.elsevier.com/retrieve/pii/S0021929017302397.

78. O'Reilly M, Whelan D, Ward T, Delahunt E, Caulfield B. Technology in strength and conditioning: Assessing bodyweight squat technique with wearable sensors. J Strength Cond Res [Internet]. 2017;31:2303–12. https://journals.lww.com/00124278-201708000-00031.

79. Shepherd JB, Giblin G, Pepping G-J, Thiel D, Rowlands D. Development and validation of a single wrist mounted inertial sensor for biomechanical performance analysis of an elite netball shot. IEEE Sens Lett [Internet]. 2017;1:1–4. http://ieeexplore.ieee.org/document/8031033/.

80. Shuai Z, Dong A, Liu H, Cui Y. Reliability and Validity of an Inertial Measurement System to Quantify Lower Extremity Joint Angle in Functional Movements. Sensors 2022, Vol 22, Page 863 [Internet]. 2022 [cited 2023 Mar 20];22:863. https://www.mdpi.com/1424-8220/22/3/863/htm.

81. Spilz A, Munz M. Automatic Assessment of Functional Movement Screening Exercises with Deep Learning Architectures. Sensors 2023, Vol 23, Page 5 [Internet]. 2022 [cited 2023 Mar 20];23:5. https://www.mdpi.com/1424-8220/23/1/5/htm.

82. Tabrizi SS, Pashazadeh S, Javani V. A deep learning approach for table tennis forehand stroke evaluation system using an IMU sensor. Versaci M, editor. Comput Intell Neurosci [Internet]. 2021;2021:1–15. https://www.hindawi.com/journals/cin/2021/5584756/.

83. Tulipani L, Boocock MG, Lomond K V., El-Gohary M, Reid DA, Henry SM. Validation of an inertial sensor system for physical therapists to quantify movement coordination during functional tasks. J Appl Biomech [Internet]. 2018;34:23–30. https://journals.humankinetics.com/view/journals/jab/34/1/article-p23.xml.

84. Xu JY, Nan X, Ebken V, Wang Y, Pottie GJ, Kaiser WJ. Integrated inertial sensors and mobile computing for real-time cycling performance guidance via pedaling profile classification. IEEE J Biomed Health Inform [Internet]. 2015;19:440–5. http://ieeexplore.ieee.org/document/6813614/.

85. Zhang J, Lockhart TE, Soangra R. Classifying lower extremity muscle fatigue during walking using machine learning and inertial sensors. Ann Biomed Eng. 2014;42:600–12. https://doi.org/10.1007/s10439-013-0917-0.

86. Kalman RE. A new approach to linear filtering and prediction problems. Journal of Basic Engineering [Internet]. 1960;82:35–45. https://asmedigitalcollection.asme.org/fluidsengineering/article/82/1/35/397706/A-New-Approach-to-Linear-Filtering-and-Prediction.

87. Zimmermann M, Sulzer W. High bandwidth orientation measurement and control based on complementary filtering. In: Proceedings of SYROCO '91, ifac symposium on robot control, Vienna. 1991.

88. Baerveldt A-J, Klang R. A low-cost and low-weight attitude estimation system for an autonomous helicopter. In: Proceedings of IEEE international conference on intelligent engineering systems. 1997. p. 391–5.

89. Zijlstra W, Hof AL. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait Posture [Internet]. 2003 [cited 2021 Jul 7];1–10. www.elsevier.com/locate/gaitpost.

90. Davey N, Wixted A, Ohgi Y, James DA. A low cost self contained platform for human motion analysis. 2008. p. 101–11. https://www.scopus.com/inward/record.uri?eid=2-s2.0-61849

091984&partnerID=40&md5=ccae96e2ecaa62d626b8f43eb ca424e5.

91. Barth AT, Hanson MA, Powell Jr. HC, Lach J. TEMPO 3.1: a body area sensor network platform for continuous movement assessment. In: 2009 sixth international workshop on wearable and implantable body sensor networks [Internet]. USA: IEEE; 2009. p. 71–6. http://ieeexplore.ieee.org/document/5226915/.

92. Cook G. Movement: functional movement systems: screening, assessment, corrective strategies. 10th ed. Aptos: On Target Publications; 2010.

93. tfa.metrics.F1Score | TensorFlow Addons [Internet]. [cited 2023 Apr 10]. https://www.tensorflow.org/addons/api_docs/python/tfa/metrics/F1Score.

94. Alzubi J, Nayyar A, Kumar A. Machine learning from theory to algorithms: an overview. J Phys Conf Ser [Internet]. 2018;1142:012012. https://doi.org/10.1088/1742-6596/1142/1/012012.

95. Fantozzi S, Giovanardi A, Magalhães FA, Di Michele R, Cortesi M, Gatta G. Assessment of three-dimensional joint kinematics of the upper limb during simulated swimming using wearable inertial-magnetic measurement units. J Sports Sci [Internet]. 2016;34:1073–80. https://doi.org/10.1080/02640414.2015.10886 59.

96. Ramsay JO, Silverman BW. Functional data analysis. 2nd ed. New York: Springer Science + Business Media, Inc.; 2005.

97. Yang C-C, Hsu Y-L. A review of accelerometry-based wearable motion detectors for physical activity monitoring. Sensors [Internet]. 2010;10:7772–88. http://www.mdpi.com/1424-8220/10/8/7772.

98. Winter DA. Biomechanics and motor control of human movement: Fourth Edition. 2009.

99. Beanland E, Main LC, Aisbett B, Gastin P, Netto K. Validation of GPS and accelerometer technology in swimming. J Sci Med Sport. 2014;17:234–8.

100. Perlmutter M, Robin L. High-performance, low cost inertial MEMS: a market in motion! In: Proceedings of the 2012 IEEE/ION position, location and navigation symposium [Internet]. IEEE; 2012. p. 225–9. http://ieeexplore.ieee.org/document/6236884/.

101. Althnian A, AlSaeed D, Al-Baity H, Samha A, Dris A Bin, Alzakari N, et al. Impact of dataset size on classification performance: an empirical evaluation in the medical domain. Applied Sciences 2021, Vol 11, Page 796 [Internet]. 2021 [cited 2023 Apr 5];11:796. https://www.mdpi.com/2076-3417/11/2/796/htm.

102. Saxena A, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, et al. A review of clustering techniques and developments. Neurocomputing. 2017;267:664–81.

103. Kotsiantis SB, Zaharakis ID, Pintelas PE. Machine learning: a review of classification and combining techniques. Artif Intell Rev. 2006;26:159–90.

104. Singh A, Thakur N, Sharma A. A review of supervised machine learning algorithms. In: 2016 3rd international conference on computing for sustainable global development (INDIACom). 2016. p. 1310–5.

105. Conceição F, Lewis M, Lopes H, Fonseca EMM. An evaluation of the accuracy and precision of jump height measurements using different technologies and analytical methods. Applied Sciences 2022, Vol 12, Page 511 [Internet]. 2022 [cited 2023 Apr 11];12:511. https://www.mdpi.com/2076-3417/12/1/511/htm.

106. Ferrero R, Gandino F, Hemmatpour M. Estimation of displacement for internet of things applications with Kalman filter. Electronics 2019, Vol 8, Page 985 [Internet]. 2019 [cited 2023 Apr 13];8:985. https://www.mdpi.com/2079-9292/8/9/985/htm.

107. Kubelka V, Reinstein M. Complementary filtering approach to orientation estimation using inertial sensors only. In: 2012 IEEE international conference on robotics and automation [Internet]. IEEE; 2012. p. 599–605. http://ieeexplore.ieee.org/document/6224564/.