Wearable Inertial Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review

Wearable Inertial Sensor Systems for Lower Limb Exercises

Martin O'Reilly [1,2], Brian Caulfield [1,2], Tomas Ward [3,4] , William Johnston[1,2] , Cailbhe Doherty [1,2]

[1] School of Public Health, Physiotherapy and Sports Science, University College Dublin, Ireland

[2] Insight Centre for Data Analytics, University College Dublin, Ireland

[3] Insight Centre for Data Analytics, Dublin City University, Ireland

[4] School of Computing, Dublin City University, Ireland

Corresponding Author Email: martin.oreilly@insight-centre.org

Corresponding Author Phone: 0035317166500

ORCID IDs:

Martin O'Reilly - 0000-0003-2425-5393

Brian Caulfield - 0000-0003-0290-9587

Tomas Ward - 0000-0002-6173-6607

William Johnston - 0000-0003-0525-6577

Cailbhe Doherty - 0000-0002-5284-856X

Abstract

Background

Analysis of lower limb exercises is traditionally completed with four distinct methods (i) 3D motion capture; (ii) depth-camera based systems (iii) visual analysis from a qualified exercise professional; (iv) self-assessment. Each method is associated with a number of limitations.

Objective

The aim of this systematic review is to synthesize and evaluate studies which have investigated the capacity for inertial measurement unit (IMU) technologies to assess movement quality in lower limb exercises.

Data Sources

A systematic review of PubMed, ScienceDirect and Scopus was conducted.

Study Eligibility Criteria

Articles written in English and published in the last 10 years which contained an IMU system for the analysis of repetition-based targeted lower limb exercises were included.

Study Appraisal and Synthesis Methods

The quality of included studies was measured using an adapted version of the STROBE assessment criteria for cross-sectional studies. The studies were categorised in to three groupings: exercise detection, movement classification or measurement validation. Each study was then qualitatively summarised.

Results

From the 2452 articles that were identified with the search strategies, 47 papers are included in this review.

Conclusions

Wearable inertial sensor systems for analysing lower limb exercises are a rapidly growing technology. Research over the past ten years has predominantly focused on validating measurements that the systems produce and classifying users' exercise quality. There have been very few user evaluation studies and no clinical trials in this field to date.

Key Points

Inertial measurement unit (IMU) systems have been extensively validated to successfully measure joint angle and temporal features during lower limb exercises.

It is less understood if IMU systems can validly compute kinetic measures pertaining to lower limb exercises.

IMU systems, which incorporate machine learning in to their data analysis pathways, have also been found to be effective in automated exercise detection and in classifying movement quality across a range of lower limb exercises.

1. Introduction

Lower limb exercises are used in rehabilitation, performance assessment, injury screening and strength and conditioning (S&C) contexts [1–3]. Movement is deemed 'optimal' during these exercises when injury-risk is minimised and performance is maximised [4]. There are currently 4 distinct methods of assessing movement during lower limb exercise: (i) 3D motion capture; (ii) depth-camera based systems (iii) visual analysis from a qualified exercise professional; (iv) self-assessment. Each method is associated with a number of limitations. For instance, 3-D motion capture systems are expensive (> €100,000) and the application of skin-mounted markers may hinder normal movement [5,6]. Furthermore, data processing can be time intensive and specific expertise is often required to interpret the processed data and to make recommendations on the observed results. Therefore, these systems are not frequently used to assess exercise technique beyond the research laboratory [7]. A cheaper and more accessible alternative (< €500) is the use of depth-camera systems such as the Microsoft Kinect. In recent times, such systems have been increasingly leveraged for both research and commercial purposes due to their low cost and ease of setup. However, such systems have several key limitations. Depth-camera systems often lack accuracy when compared to gold-standard marker-based systems. Such systems operate by tracking specific body locations and re-creating body segments based on these locations. As such, confusion and resultant poor accuracy is often caused by crossing over of body segments, unsuitable lighting (outdoors), movement of clothing and movement of other people [8]. As a result, users often must engage in time intensive manual re-labelling of body segments to ensure an accurate system. Secondly, while these systems are relatively unobtrusive, they do require the user to set up a camera in an empty $2m^2$ area. However, depending on the application space (clinic or gym), this may not be possible due to the presence of other people and equipment (squat rack/ weight bench) that may confuse the system, resulting in poor accuracy. In clinical and gym-based settings, visual assessment is typically used to assess lower limb exercises. Visual assessment of human biomechanics is subjective and unreliable amongst novices and experts alike, as the need to visually assess numerous constituent components simultaneously is challenging [9]. This issue is compounded by the fact that athletes/clients may not be able to afford the supervision of a qualified professional (such as a physiotherapist, athletic therapist or personal trainer) in many instances. For this reason, individuals largely rely on self-assessment of their exercise technique in gym-based settings. The obvious limitation with this approach is that the individual may lack the knowledge required to assess their movement patterns, while simultaneously completing an exertive movement and assessing it without bias can be difficult [10].

Due to these limitations, in the past 15 years there has been an increase in interest in the automated assessment of lower limb exercises with wearable inertial measurement units (IMUs). Wearable IMUs are small, inexpensive sensing units (≈ €50-1,000) that consist of accelerometers, gyroscopes and/or magnetometers. They are able to acquire data pertaining to the inertial motion and 3D orientation of individual limb segments [11,12]. Self-contained, wireless IMU devices are easy to set up, and allow for the acquisition of human movement data in unconstrained environments [13]. IMU systems can robustly track a variety of postures in the complex environment associated with training in the 'real-world', unlike camera-based systems, which are hampered by location, occlusion and lighting issues [14]. IMUs have also been shown to be as effective as marker-based systems at measuring joint angles [7,15,16]. Therefore, IMUs have been recently employed for analysing a range of components of lower limb exercises. This includes detecting and quantifying the number of repetitions that are completed of a given exercise [17,18] , computing the range of motion (ROM) at key joints during these repetitions [19,20], temporal analysis of exercises [21,22], classifying one's performance of an exercise as acceptable or as a specific deviation from acceptable [3,23], or extracting exercise performance measures such as jump height and reactive strength index [24].

In the past decade a number of reviews have assessed the literature pertaining to exercise analysis with wearable sensors. Fong and Chan reviewed the use of wearable IMUs in lower limb biomechanics studies, however the focus of this work was broad, and predominantly reviewed gait based papers [25]. Another early review covered the broad scope of health and wellness, rehabilitation, and injury prevention with both wearable and ambient sensor systems [26]. The field has expanded considerably since then. Recently, a systematic review was published by Wang et al. which classified studies involving upper limb wearable systems for rehabilitation [27]. The 'wearability' of such systems and evidence supporting the systems' effectiveness were also reviewed. Prior to this, this group published a review of studies on upper limb rehabilitation systems from 2008 to 2013 [28]. A variety of works have given an in depth summary of movement measurement and analysis technologies, however these do not focus on exercise analysis or the lower limb [29–31]. Cuesta-Vargas et al. reviewed the use of inertial sensors in human motion analysis and showed their capability for task-specific analysis [32]. Other studies have investigated how feedback affects therapy outcomes, however these systems did not necessarily involve wearable IMUs and focused predominantly on the upper extremity [33–35]. To date, a contemporary systematic review investigating the capacity for IMU technologies to quantify movement quality during lower limb exercises is not available. Therefore, the aim of this systematic review is to synthesize and evaluate studies which have investigated the capacity for IMU technologies to assess movement quality in lower limb exercises such as straight-leg raises, squats and countermovement jumps. In particular, we aim to describe the sensing set-ups used, inclusive of type (accelerometer and/or gyroscope and/or magnetometer), number and position of the sensing units. We also aim to describe the measurements each system extracted from the sensing units (e.g. ROM, power) and how they were validated. We will also establish which exercises were analysed by such systems. This review serves to summarise a rapidly growing field which has not been specifically reviewed in over 7 years [25]. It will identify clear gaps in the literature which are of interest to the research community and can be used as a resource for sports-medicine practitioners to build an understanding of the capabilities of IMU systems in assessing lower limb exercises. We hypothesise that IMU systems may be an effective and affordable tool to analyse components of lower limb exercises objectively and efficiently.

2. Methods

2.1. Literature Search Strategy and Study Selection Process

The protocol for this review was performed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [36]. A literature search was completed within the following 3 databases: PubMed, Scopus and ScienceDirect. Papers regarding the following were selected: exercise, lower body, movement monitoring and IMUs. MeSh (Medical Subject Heading) terms or title/abstract keywords and their synonyms and spelling variations were used in several combinations and modified for every database. Articles published from January 2007 to May 2017 were reviewed. The 2007 start date was chosen, to minimise irrelevant search results, as it represents the first known paper published in the field [17]. The general search strategy including the search terms used, are listed in Table 1. This search includes refereed journal papers and peer-reviewed articles published in conference proceedings. Only articles written in the English language were included. The article selection process consisted of the following steps using the PRISMA [36] guidelines (Figure 1): 1) A computerized search strategy was performed for the period January 2007 until September 2017; 2) After removal of duplicates, titles and abstracts of the remaining articles were screened; 3) The reviewer read the full texts and selected articles based on the inclusion/exclusion criteria (Table 2). In cases where a journal paper covered the contents reported in the earlier conference publications, the journal paper was preferred over the conference paper. In cases where the overlap was only partial, multiple publications were used as sources. Due to the relative novelty of IMU technologies, the grey literature was not searched; only peer-reviewed scientific articles were eligible for inclusion. We deemed this appropriate due to the non-interventional nature of studies in this field.

Table 1 about here

Table 2 about here

2.2. Data Extraction Process

Data extraction was completed by two authors (MOR and CD). Where discrepancies occurred, these were discussed and the associated papers were reassessed. A standardised data extraction form was utilised. Details about the study design, the exercises investigated, the sensor systems (e.g. accelerometer-only vs accelerometer + gyroscope) and the set-ups (e.g. multi-site vs single-site) used were ascertained. The studies were divided into three categories based on the aims/objectives of this review: exercise detection (ED); movement classification (MC); measurement validation (MV). Each study was then qualitatively summarised (aims, findings and conclusions based on these findings).

2.3. Assessment of Study Quality

Two authors (MOR and CD) evaluated the quality of the studies deemed eligible for inclusion using an adapted version of the STROBE assessment criteria for cross-sectional studies [37], which was devised by author consensus. Specifically, each study was rated on 10 specific criteria which were derived from items 1, 3, 6, 8, 11, 14, 18, 19, 20 and 22 of the original checklist. In cases where the authors completing paper rating (MOR or CD) were an author of a paper included in this review, the paper was instead rated by a different author of this paper (WJ) to minimise the risk of bias. Final study ratings for each reviewer were collated and examined for discrepancies. Any inter-rater disagreement was resolved by consensus decision. Once consensus was reached for all study ratings, overall quality scores were collated by summing those criteria, providing a score out of 10. Studies were considered to be of high quality when >7 domains were scored as high (1). If >3 domains were scored as low (0), the study was considered of low quality.

3. Results

3.1. Database Search and Paper Lists

An overview of the results in the different stages of the article selection process is shown in Figure 1. From the 2452 articles that were identified with the search strategies, 47 papers are included in this review following the selection process.

The quality of the included reviews is displayed in Table 3. Based on our pre-defined criteria, 26 of the 47 included studies were deemed as being of high quality. Briefly, most studies adequately reported the methods of data acquisition (42/47), the outcome variables of interest and the method of statistical analysis employed (43/47). In contrast, many authors did not adequately discuss the limitations of the study (24/47), detail the eligibility criteria of the included sample (19/47) or cite relevant literature when discussing their results (27/47).

Table 3 about here

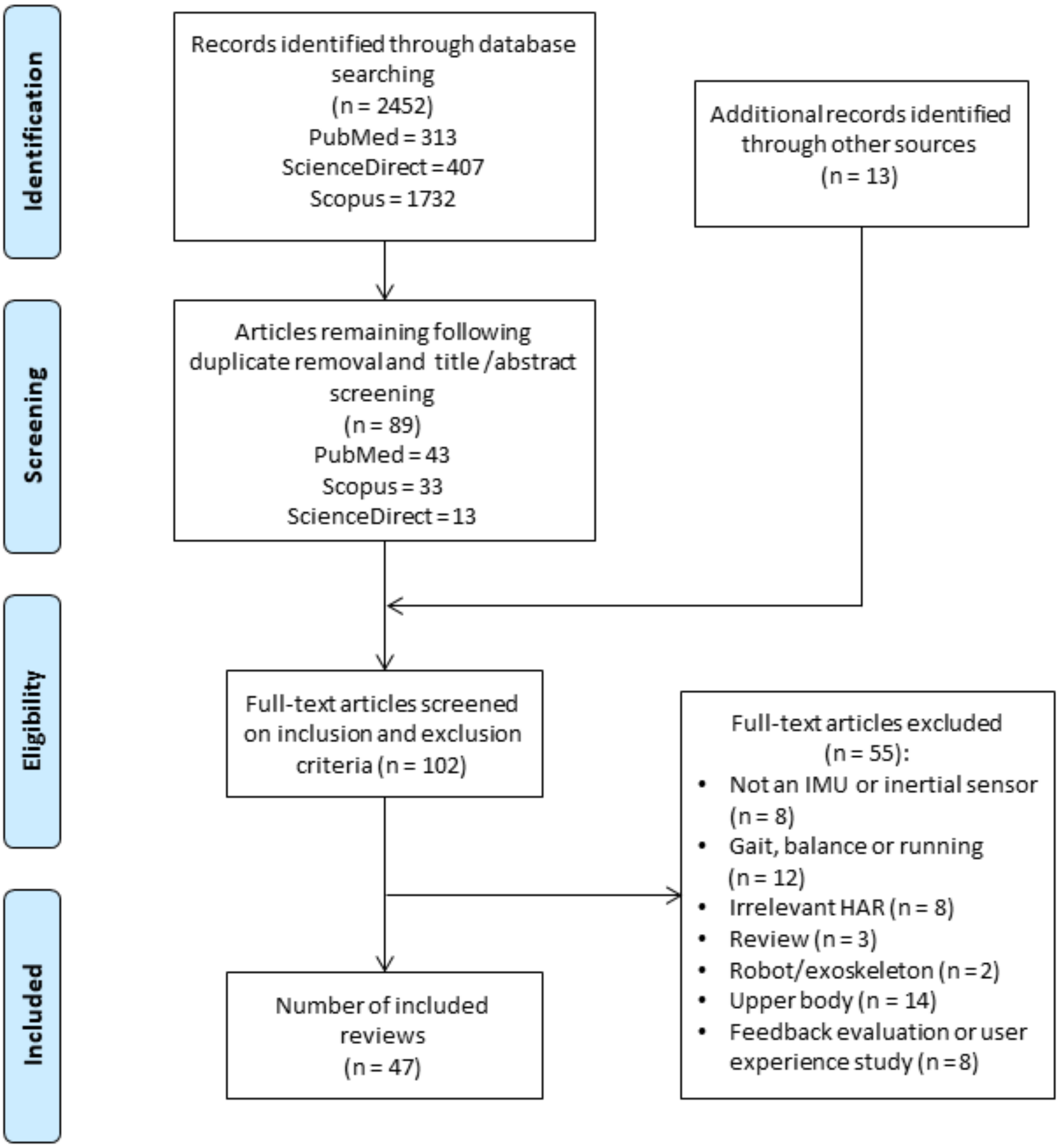

Figure 1: Prisma flowchart of the results from the literature search.

Figure legend: IMU = inertial measurement unit, HAR = human activity recognition

3.2. Sensor Set-ups

Table 4 categorizes the included articles based on whether the systems they adopted used multiple/single sensor units, compared sensor units at a variety of anatomical locations, and/or compared multiple sensor set-ups to single sensor set-ups for each application.

There was a large degree of heterogeneity in the included studies' sensor set-ups. In particular, the types of sensors on board each sensing unit (accelerometer and/or gyroscope and/or magnetometer) and the number of sensing units required to be worn by system users varied. Table 5 demonstrates the distribution of sensors used in the included studies.

Table 4 and 5 about here

3.3. Exercises Investigated Versus Study design

In the included studies, a total of fifty-three exercises were evaluated using a wearable inertial sensor system (Table 6). The most commonly investigated single-joint, uni-planar exercise was the lying straight leg-raise. There were three single-joint multi-planar exercises investigated. There were also two multi-joint, uni-planar exercises and twenty-six multi-joint, multi-planar exercises investigated. The most investigated of these were the sit-to-stand and squat exercises.

Table 6 about here

3.4. Qualitative Review

3.4.1. Measurement Validation

Twenty-eight studies identified for inclusion in this review attempted to validate wearable motion sensor systems [7,20–22,24,38–61]. These twenty-eight studies were categorised as evaluating either concurrent validity (Table 7) or construct validity (Table 8). For the purposes of this review, concurrent validity was defined as when a newly developed tool such as a wearable sensor system is compared to another test which is considered to be the "gold standard" to measure the construct in question [62]. Construct validity compares a new wearable system's output to another test that measures a similar construct but that is not a "gold standard" (convergent validity), or evaluates the system's capacity to discriminate between known-groups in a cross sectional (discriminative validity; known groups) or longitudinal (discriminative validity; responsiveness) manner [62].

Concurrent Validity

Seventeen of the studies included in this review sought to compare a wearable sensor system's output to a tool used in current clinical practice (e.g. goniometer for joint angle measurement) or gold standard biomechanical measurement tools (e.g. optoelectronic motion capture systems and force plates) [7,20,21,24,38–50]. These studies are summarised in Table 7.

Table 7 about here

Construct validity

Eleven studies investigated the construct validity of wearable motion systems for specific applications in tracking lower limb exercises. Of these, four assessed convergent validity [51–54]. Five studies pertained to known-groups validity [22,55–58]. Two studies evaluated the longitudinal validity of a lower limb wearable sensor system in assessing joint ROM throughout a rehabilitation programme [60,61]. All eleven studies which predominantly evaluated construct validity are summarised in Table 8.

Table 8 about here

3.4.2. Exercise Detection

Ten studies were identified for which automated detection of the exercise being completed was a key objective [14,17–19,63–68]. These studies are summarised in Table 9. It is difficult to directly compare the exercise detection sensitivity, specificity and accuracy across different studies. This is due to the vastly different data sets and cross-validation method used to compute system accuracy.

Table 9 about here

3.4.3. Movement Classification

Eleven studies investigated the utilisation of wearable IMU systems for quantifying exercise technique [3,19,23,64,69–75]. Table 10 summarises the sensing set-ups, movement measure which was classified, methodology and performance metrics for each system identified in this area.

Table 10 about here

4. Discussion

4.1. Sensor Set-ups

Various approaches have been employed when considering sensor set-ups for analysing lower limb exercises. As shown in Table 6, some studies opted solely for using one sensor type (accelerometer, gyroscope or magnetometer) whereas others opted for combinations of these in IMUs. The use of additional on-board sensors will reduce a sensing unit's battery life, but allows for a greater variety of motion data to be captured from a user [12]. Combining accelerometer, gyroscope and magnetometer data also allows for improved accuracy in computing each unit's 3D orientation [11]. The authors of this review believe that collecting data with all three inertial sensor types and then comparing system quality (i.e. accuracy or agreement with gold standard measurement) with individual sensor types or reduced combinations is the best approach. This enables a systematic approach to assess the cost-benefit of using additional sensors on board each IMU.

Similarly, the cost-benefit of using wearable sensing set-ups which use multiple sensing units can be compared to reduced sensor sets or single sensing unit set-ups by initially completing data collection with comprehensive set-ups. System efficacy can then be assessed when using data from multiple sensing units and each reduced combination of sensing units. This approach has been applied in movement classification and has shown promising results for single sensor systems in analysing early stage and late stage, lower limb rehabilitation exercises [3,70,71]. Using reduced sensing set-ups potentially reduces the total cost of systems and increases their practicality for end users.

A criticism of the approach to sensor set-up employed by all the included studies in this review is that none reported evidence of engaging target end-users when selecting their set-up. As previously mentioned, sensor set-up can influence a system's cost, usability, battery-life, accuracy and functionality. It is likely that the importance of each of these factors would vary across differing types of end-users. For instance, a recreational gym goer may prioritise cost, usability and battery life which may favour a minimal sensor set-up whereas an elite sports team may prioritise accuracy and functionality which could favour a comprehensive set-up. It is the authors' contention that completing relevant qualitative research regarding sensor set-up with the target end-users of a system, in advance of its development, could be of great benefit. It could also streamline the process of transferring systems from a research environment in to a real-world offering.

4.2. Measurement Validation Studies

The concurrent validity of wearable motion sensor based systems to extract useful measures in lower limb exercises has been well investigated. The articles included in this review demonstrate these systems' validity in measuring joint angle and ROM, in a broad variety of exercises ranging from sit-to-stands [48,49] to lunges [42]. It is of note that a large proportion of these studies have high paper quality ratings (Table 3). They also used optoelectronic motion capture, a biomechanical research gold-standard measurement tool [20,39,50] or goniometry, a clinical gold-standard [41–43] as the measurement comparator, which adds to the strength of this literature. Therefore, the evidence that IMU systems can validly measure joint angle and ROM is strong. However, for applications requiring maximal accuracy in these measures, an interesting area of further research may be identifying the sensing-unit placement position, for various body segments and exercises, which optimally agrees with an optoelectronic motion capture system. With the exception of Faber et al. [40], this is a widely under investigated field. Research has also demonstrated the concurrent validity of wearable motion sensor based system measures, with force plate and optoelectronic motion capture data, to compute temporal features of exercises. A number of these studies have high quality ratings (Table 3) and analyse exercises ranging from five time sit-to-stand tests [48], to deadlifts [21] to drop jumps [24,44]. It is less understood if wearable motion sensor systems are useful in estimating kinetic measures such as peak vertical force and power during exercises. With the exception of Zijlstra et al.'s study on vertical power during the sit to stand exercise [49] the 'high' quality work to date, in this area, has shown these measures have a lower agreement with gold-standard biomechanical measurement systems than joint angle or temporal features [44,46,48]. Further research is required to investigate if the results are unique to different types of exercises or if they can be improved through employing different signal processing techniques.

The construct validity of wearable motion systems for a range of applications has also been well demonstrated at this point. Longitudinal validity has been shown via progress tracking in ROM through a rehabilitation programme [59,60]. Known-groups validity has been demonstrated through capturing different movement profiles in specific exercises between injured and non-injured individuals [56,57]. However, it should be noted these studies involved very small samples. Fitzgerald et al. [56] compared just a single injured and non-injured participant and Ai et al. [57] compared 3 non-injured participants with 1 participant with polymyositis and 1 participant with lower back pain. Both papers also received a low paper quality rating (Table 3). Therefore, it is currently difficult to conclude that IMU systems can differentiate movement profiles from injured and non-injured groups and that the presented results are not just due to chance. Future validation studies of this fashion would benefit from employing larger participant groups of both injured and non-injured individuals. This would allow for statistical analyses comparing the groups. The results could be further strengthened by demonstrating concurrent validity of the reported measurements with an existing gold-standard measurement tool.

Whilst measurement validation studies are the most researched category in this field, there is still scope for more investigation. It is the authors' contention that studies which validate temporal feature, joint angle and ROM measurements, computed from multiple sensing units, in lower limb exercises are likely to produce favourable results, but produce little new knowledge for the field. However, creating predictive algorithms for kinetic measurements, such as that in Setuain et al. [46] and Zijlstra et al. [49], or joint angle estimations with a single IMU in [7,43] are still widely under-investigated areas with much room for advancement. Research which employs larger samples and assesses known-groups validity between injured and non-injured participants could also progress this area of the field.

4.3. Exercise Detection Systems

Ten studies included in this survey have demonstrated the efficacy of wearable inertial sensor systems to automatically identify the exercises being completed by users [14,17–19,63–68]. Exercise classification may serve as a useful input to an automated exercise tracking system or automated exercise logbook system. All studies demonstrated that a machine learning based classification approach is an effective data analysis approach for this task. With this in mind, it is interesting to note that only four studies in this area involved more than twenty participants [18,66–68]. As with many classification problems, exercise detection results may be improved by collecting larger data sets from more participants [76]. Future work in this area could also consider the practicality of the system for end users. Utilising a single IMU system for exercise detection [14,18,66–68], may be most desirable for end users. This will also reduce the cost of the exercise detection system. A potential methodological flaw in some of the reviewed studies is the inclusion of repetitions of exercises from the same participant in both training and test data. This can produce unrealistically high accuracy scores following cross-validation on a data set. This may have happened in a number of studies which used cross-validation methods such as leave-one-out-cross-validation (LOOCV), repeated random sub-sampling (RRSS) and K-fold cross-validation where each fold did not represent all of one participant's data [14,19,63,64]. Future work may also benefit from utilising deep learning techniques for classification such as the convolutional neural networks approach demonstrated by Veiga et al. [68]. Such classification methodologies have recently been shown to have many benefits when compared to traditional machine learning classification techniques when analysing time-series data, including reducing the risk of overfitting and improving system accuracy [77,78]. The method of Veiga et al. which uses a machine vision approach within a deep learning context also has the advantage of allowing interpretability of classification based on the visual appearance of the time series.

4.4. Movement Classification Systems

Despite having first being investigated in 2010 [23], lower limb, wearable, movement classification systems are still a relatively under investigated area. Five of the eleven published papers are relatively small-scale, with ten or fewer participants [19,23,64,69,75]. Movement classification systems have the potential to augment current clinical practice, providing users with feedback relating to their exercise technique in an unsupervised setting [79]. Most of the published work on movement classification pertains to a limited number of rehabilitation and S&C exercises (Table 10) [3,19,23,64,70–72,74]. There is therefore potential to investigate movement classification with larger data sets and across a range of other exercises. This work could also compare binary and multi-label classification techniques and comprehensive and minimal sensing set-ups for such exercises, as in the work conducted by Giggins and colleagues [3]. Such work could also compare a variety of classification strategies (e.g. random forests [80], support vector machines [81] , k nearest neighbours [82]) and should use appropriate cross validation techniques to estimate system efficacy (accuracy, sensitivity and specificity). The studies which used leave-one-subject-out-cross-validation (LOSOCV) to validate their global movement classification system report that this is the most appropriate cross validation method to estimate the efficacy of the system for a new user who is not included in the classifier's training data [3,23]. They also warn that including exercise repetitions from the same participant in both training and test data, as in Yurtman and Barshan [19] and Whelan et al. [70], can produce artificially high efficacy scores which do not transfer to real-world systems. Only one study experimentally evaluated the real-world accuracy of a movement classification system [75], which is recommended for future studies where possible as it negates the limitations of cross-validation techniques in assessing system efficacy. The efficacy scores presented by authors included accuracy, sensitivity/precision/true positive rate and specificity/recall/true negative rate. It is best to provide as many metrics as possible to allow a reader to understand a classification system's strengths and weaknesses.

The general data analysis approach for all studies appears to be first completing signal pre-processing, signal segmentation, computing features from the signals and placing them in feature vectors which will be used to train and evaluate different classification algorithms [79]. Some recent studies also compare the effectiveness of global and personalised classification systems [73,74], whereby a personalised classification system is one which is trained from data from an individual and developed specifically for this individual and a global classification system is trained with data from many individuals and can be used by individuals not included in the training data. Analysis has shown that the personalised systems are more computationally efficient and accurate than global ones [73–75], which has enabled movement classification systems to be developed with single sensing unit rather than multiple sensing units. However, the time required to collect data from and train a classification system for every individual who requires a movement classification system is a significant practical limitation which may hinder the uptake of movement classification systems in clinical practice. Creating tools which streamline this process, as in O'Reilly et al. [75], could be an important avenue of research.

4.5. Review Limitations

Despite the strengths of this systematic review, it is important to consider several limitations when interpreting the results. Studies were not included if they were not published in the English language, which may influence the outcomes of our analyses, despite the probability that authors of high quality surveys would aim for publication in high-impact journals published in the English language in the pursuit of superior dissemination of output data. Additionally, while the data extraction and paper rating were both completed by two authors (MOR and CD), the initial search, title and abstract screening was only completed by one author (MOR). Therefore, the possibility of erroneous exclusion of a valid study from this review exists. Finally, the protocol for this review was not registered prior to its completion. We deemed pre-registration inappropriate due to the expected heterogeneity in the evidence base. The scientific field of movement detection, classification and feedback using IMU is a burgeoning area that spans a multitude of disciplines, making systematic review and curation using a predefined protocol difficult.

4.6. Practical Implications

The practical implications of these findings deserve consideration. For members of the public or sports-medicine practitioners considering purchase of a commercially available sensing system to support lower-limb exercise completion for themselves or their patient/client, we would recommend a three-step appraisal process to determine the utility of the system under consideration: First, for what purpose is the system designed? Whether the system in question is for exercise detection (e.g. repetition (rep) counting during a deadlift) or classification (e.g. aberrant form during a squat) should be determined. Second, the system's validity for the stated purpose in the same population should be confirmed. A system that is valid for detection is not commensurate with a system that is valid for classification. Similarly, the validity of a system tested on a healthy population does not necessarily extend to a pathological population. Finally, at what cost does the feature-set come? More expensive systems which incorporate a greater number of sensing units are likely to be more accurate, but the benefit of this increased accuracy should be identified. For instance, sports-medicine practitioners may seek higher accuracies for a system they plan to deploy in a cohort who are at a higher risk of injury to detect aberrant exercise technique than for a cohort in whom they are seeking to quantify exercise load through rep-counting. As such, it may be more appropriate to leverage a more accurate and expensive multi-sensor system in the aberrant exercise cohort than in the rep counting cohort.

The number of commercially available sensing systems for exercise detection, classification and feedback will likely increase in the coming years, and the methods underlying these systems will evolve as the field of research progresses. We believe the appraisal process outlined above will accommodate new developments in the field.

This systematic review has also led to a number of recommendations for researchers developing wearable motion sensor systems for analysing lower limb exercises. A summary of such recommendations can be found in Table 11.

5. Conclusion

Wearable inertial sensor systems for analysing lower limb exercises is a rapidly growing technology. Research over the past ten years has involved both the development and evaluation of such systems. The research to date has predominantly focused on validating measurements that the systems produce and classifying technique quality in the exercises (Tables 7, 8 and 10). A smaller number of studies have evaluated the ability of the systems to detect exercise type. Table 6 shows the fifty-three exercises which have currently been incorporated into such systems and highlights gaps in the literature which warrant further research. One such gap is that there are a limited number of studies which classify movement quality in jumping exercises. There exist a vast amount of considerations for future research in this field as outlined in Table 11. Moreover, there have been very few user evaluation studies and no clinical trials evaluating wearable inertial sensor systems for lower limb exercises. Such studies will be essential in producing knowledge which will catalyse the movement of these systems from laboratory based studies in to real world applications for sports-medicine practitioners and people completing lower limb exercises.

Compliance with Ethical Standards

Funding

Martin O'Reilly was partially funded for this work by the Irish Research Council as part of a Postgraduate Enterprise Partnership Scheme with Shimmer (EPSPG/2013/574). Brian Caulfield, Tomas Ward, William Johnston and Cailbhe Doherty were funded and Martin O'Reilly was partially funded by Science Foundation Ireland under their grant for the Insight Centre for Data Analytics (SFI/12/RC/2289). These funding bodies had no influence on the data collection, data analysis, data interpretation or approval/disapproval of publication.

Conflicts of Interest

Martin O'Reilly, Brian Caulfield, Tomas Ward, William Johnston and Cailbhe Doherty declare that they have no conflicts of interest relevant to the content of this review.

References

1. Waller M, Gersick M, Holman D. Various jump training styles for improvement of vertical jump performance. Strength Cond J. (Lippincott Williams Wilkins) [Internet]. 2013;35:82–9. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=85709355&lang=pt-br&site=ehost-live

2. Whittaker JL, Booysen N, de la Motte S, Dennett L, Lewis CL, Wilson D. Predicting sport and occupational lower extremity injury risk through movement quality screening: a systematic review. Br J Sports Med. [Internet]. 2017;51:580–5. Available from: http://bjsm.bmj.com/lookup/doi/10.1136/bjsports-2016-096760

3. Giggins OM, Sweeney KT, Caulfield B. Rehabilitation exercise assessment using inertial sensors: a cross-sectional analytical study. J Neuroeng Rehabil. [Internet]. 2014;11:158. Available from: http://www.jneuroengrehab.com/content/11/1/158

4. Farrokhi S, Pollard CD, Souza RB, Chen Y-J, Reischl S, Powers CM. Trunk position influences the kinematics, kinetics, and muscle activity of the lead lower extremity during the forward lunge exercise. J Orthop Sports Phys Ther. 2008;38:403–9.

5. Ahmadi A, Mitchell E, Destelle F, Gowing M, O'Connor NE, Richter C. Automatic activity classification and movement assessment during a sports training session using wearable inertial sensors. Proc11th Int Conf Wearable Implant Body Sens Networks, BSN 2014. 2014. p. 98–103.

6. Bonnechère B, Jansen B, Salvia P, Bouzahouene H, Omelina L, Moiseev F. Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry. Gait Posture. 2014;39:593–8.

7. Bonnet V, Mazza C, Fraisse P, Cappozzo A. An optimization algorithm for joint mechanics estimate using inertial measurement unit data during a squat task. Proc Annu Int Conf IEEE Eng Med Biol Soc, EMBS. 2011. p. 3488–91.

8. Arulampalam J, Pierrepont J, Kark L. Markerless motion capture: validity of microsoft kinect cameras and ipisoft. Gait Posture [Internet]. Elsevier; 2018;42:S76. Available from: http://dx.doi.org/10.1016/j.gaitpost.2015.06.141

9. Whiteside D, Deneweth JM, Pohorence MA, Sandoval B, Russell JR, McLean SG. Grading the functional movement screen™: A comparison of manual (real-time) and objective methods. [Internet]. J Strength Cond Res. 2014. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25162646

10. John OP, Robins RW, Craik KH, Dawes RM, Funder DC, Kemis M. Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. J Pers Soc Psychol. 1994;66:206–19.

11. Madgwick SOH, Harrison AJL, Vaidyanathan R. Estimation of IMU and MARG orientation using a gradient descent algorithm. IEEE Int Conf Rehabil Robot. 2011.

12. Burns A, Greene BR, McGrath MJ, O'Shea TJ, Kuris B, Ayer SM. SHIMMER™ – A wireless sensor platform for noninvasive biomedical research. IEEE Sens J. [Internet]. 2010;10:1527–34. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5482214

13. McGrath D, Greene BR, O'Donovan KJ, Caulfield B. Gyroscope-based assessment of temporal gait parameters during treadmill walking and running. Sport Eng. 2012;15:207–13.

14. Morris D, Saponas TS, Guillory A, Kelner I. RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises. CHI '14 Proc SIGCHI Conf Hum Factors Comput Syst. [Internet]. ACM; 2014. p. 3225–34. Available from: http://dl.acm.org/citation.cfm?id=2557116

15. Leardini A, Lullini G, Giannini S, Berti L, Ortolani M, Caravaggi P. Validation of the angular measurements of a new inertial-measurement-unit based rehabilitation system: comparison with state-of-the-art gait analysis. J Neuroeng Rehabil. 2014;11:1–7.

16. Tang Z, Sekine M, Tamura T, Tanaka N, Yoshida M, Chen W. Measurement and estimation of 3D orientation using magnetic and inertial sensors. Adv Biomed Eng. 2015;4:135–43.

17. Chang K, Chen MY, Canny J. Tracking free-weight exercises. Ubiquitous Comput. 2007. p. 19–37.

18. Giggins O, Sweeney KT, Caulfield B. The use of inertial sensors for the classification of rehabilitation exercises. Conf Proc IEEE Eng Med Biol Soc. 2014;2965–8.

19. Yurtman A, Barshan B. Automated evaluation of physical therapy exercises using multi-template dynamic time warping on wearable sensor signals. Comput Methods Programs Biomed. [Internet]. 2014;117:189–207. Available from: http://www.sciencedirect.com/science/article/pii/S0169260714002910

20. Lin JFS, Kulić D. Human pose recovery using wireless inertial measurement units. Physiol Meas. [Internet]. 2012;33:2099–115. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23174667

21. Gleadhill S, Lee JB, James D. The development and validation of using inertial sensors to monitor postural change in resistance exercise. J Biomech. [Internet]. 2016;49:1259–63. Available from: http://www.sciencedirect.com/science/article/pii/S0021929016302937

22. Setuain I, González-Izal M, Alfaro J, Gorostiaga E, Izquierdo M. Acceleration and orientation jumping performance differences among elite professional male handball players with or without previous ACL reconstruction: an inertial sensor unit-based study. PM&R [Internet]. 2015;7:1243–53. Available from: http://www.sciencedirect.com/science/article/pii/S1934148215002567

23. Taylor PE, Almeida GJM, Kanade T, Hodgins JK. Classifying human motion quality for knee osteoarthritis using accelerometers. 2010 Annu Int Conf IEEE Eng Med Biol Soc. 2010. p. 339–43.

24. Patterson M, Caulfield B. A method for monitoring reactive strength index. Procedia Eng. [Internet]. 2010;2:3115–20. Available from: http://dx.doi.org/10.1016/j.proeng.2010.04.120

25. Fong DTP, Chan YY. The use of wearable inertial motion sensors in human lower limb biomechanics studies: A systematic review. Sensors (Switzerland). 2010. p. 11556–65.

26. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil. [Internet]. 2012;9:21. Available from: http://www.jneuroengrehab.com/content/9/1/21

27. Wang Q, Markopoulos P, Yu B, Chen W, Timmermans A. Interactive wearable systems for upper body rehabilitation: a systematic review. J Neuroeng Rehabil. [Internet]. London: BioMed Central; 2017;14:20. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5346195/

28. Wang Q, Chen W, Markopoulos P. Literature review on wearable systems in upper extremity rehabilitation. Biomed Heal Informatics (BHI), 2014 IEEE-EMBS Int Conf IEEE; 2014. p. 551–5.

29. Wong WY, Wong MS, Lo KH. Clinical applications of sensors for human posture and movement analysis: a review. Prosthet Orthot Int. [Internet]. SAGE Publications; 2007;31:62–75. Available from: http://dx.doi.org/10.1080/03093640600983949

30. Noorkõiv M, Rodgers H, Price CI. Accelerometer measurement of upper extremity movement after stroke: a systematic review of clinical studies. J Neuroeng Rehabil. [Internet]. London: BioMed Central; 2014;11:144. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4197318/

31. Howcroft J, Kofman J, Lemaire ED. Review of fall risk assessment in geriatric populations using inertial sensors. J Neuroeng Rehabil. [Internet]. BioMed Central; 2013;10:91. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3751184/

32. Cuesta-Vargas AI, Galán-Mercant A, Williams JM. The use of inertial sensors system for human motion analysis. Phys Ther Rev. [Internet]. 2010;15:462–73. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3566464/

33. Ribeiro DC, Sole G, Abbott JH, Milosavljevic S. Extrinsic feedback and management of low back pain: A critical review of the literature. Man Ther. [Internet]. Elsevier; 2017;16:231–9. Available from:

http://dx.doi.org/10.1016/j.math.2010.12.001

34. Subramanian SK, Massie CL, Malcolm MP, Levin MF. Does provision of extrinsic feedback result in improved motor learning in the upper limb poststroke? A systematic review of the evidence. Neurorehabil Neural Repair [Internet]. SAGE Publications; 2009;24:113–24. Available from: http://dx.doi.org/10.1177/1545968309349941

35. Saposnik G, Levin M. Virtual reality in stroke rehabilitation. Stroke [Internet]. 2011;42:1380 LP-1386. Available from: http://stroke.ahajournals.org/content/42/5/1380.abstract

36. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement. PLoS Med. 2009:6(7):e1000097.

37. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61:344–9.

38. Haladjian J, Hodaie Z, Xu H, Yigin M, Bruegge B, Fink M. KneeHapp: a bandage for rehabilitation of knee injuries. Proc 2015 ACM Int Symp Wearable Comput. 2015;181–4.

39. Bolink SAAN, Naisas H, Senden R, Essers H, Heyligers IC, Meijer K. Validity of an inertial measurement unit to assess pelvic orientation angles during gait, sit–stand transfers and step-up transfers: Comparison with an optoelectronic motion capture system. Med Eng Phys. [Internet]. 2016;38:225–31. Available from: http://www.sciencedirect.com/science/article/pii/S1350453315002672

40. Faber GS, Chang CC, Kingma I, Dennerlein JT. Lifting style and participant's sex do not affect optimal inertial sensor location for ambulatory assessment of trunk inclination. J Biomech. [Internet]. Elsevier; 2013;46:1027–30. Available from: http://dx.doi.org/10.1016/j.jbiomech.2012.12.007

41. Mehta SP, Barker K, Bowman B, Galloway H, Oliashirazi N, Oliashirazi A. Reliability, concurrent validity, and minimal detectable change for iphone goniometer app in assessing knee range of motion. J Knee Surg. Thieme Medical Publishers; 2016;

42. Romero Morales C, Calvo Lobo C, Rodríguez Sanz D, Sanz Corbalán I, Ruiz Ruiz B, López López D. The concurrent validity and reliability of the Leg Motion system for measuring ankle dorsiflexion range of motion in older adults. PeerJ [Internet]. San Francisco, USA: PeerJ Inc.; 2017;5:e2820. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5214953/

43. Bonnet V, Mazza C, Fraisse P, Cappozzo A. Real-time estimate of body kinematics during a planar squat task using a single inertial measurement unit. IEEE Trans Biomed Eng. IEEE; 2013;60:1920–6.

44. Quagliarella L, Sasanelli N, Belgiovine G, Moretti L, Moretti B. Evaluation of standing vertical jump by ankles acceleration measurement. J Strength Cond Res. LWW; 2010;24:1229–36.

45. Gordon D, Mullane SL, Conway PP, West AA. Development of a novel system for monitoring strength and conditioning in elite athletes. Procedia Eng. [Internet]. 2012;34:496–501. Available from: http://www.sciencedirect.com/science/article/pii/S1877705812016980

46. Setuain I, Martinikorena J, Gonzalez-Izal M, Martinez-Ramirez a., Gómez M, Alfaro-Adrián J. Vertical jumping biomechanical evaluation through the use of an inertial sensor-based technology. J Sports Sci. [Internet]. Routledge; 2016;34:843–51. Available from: http://www.tandfonline.com/doi/full/10.1080/02640414.2015.1075057

47. Rawson ES, Walsh TM. Estimation of resistance exercise energy expenditure using accelerometry. Med Sci. Sports Exerc. [Internet]. 2010 [cited 2013 Dec 20];42:622–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19952824

48. Papi E, Osei-Kuffour D, Chen Y-MA, McGregor AH. Use of wearable technology for performance assessment: A validation study. Med Eng Phys. [Internet]. Butterworth-Heinemann; 2015;37:698–704. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4510317/

49. Zijlstra W, Bisseling RW, Schlumbohm S, Baldus H. A body-fixed-sensor-based analysis of power during

sit-to-stand movements. Gait Posture [Internet]. 2010;31:272–8. Available from: http://www.sciencedirect.com/science/article/pii/S0966636209006559

50. Charlton PC, Mentiplay BF, Pua YH, Clark RA. Reliability and concurrent validity of a Smartphone, bubble inclinometer and motion analysis system for measurement of hip joint range of motion. J Sci Med Sport [Internet]. Sports Medicine Australia; 2015;18:262–7. Available from: http://dx.doi.org/10.1016/j.jsams.2014.04.008

51. Bo APL, Hayashibe M, Poignet P. Joint angle estimation in rehabilitation with inertial sensors and its integration with kinect. Eng Med Biol Soc EMBC, 2011 Annu Int Conf IEEE. IEEE; 2011. p. 3479–83.

52. Pernek I, Hummel KA, Kokol P. Exercise repetition detection for resistance training based on smartphones. Pers Ubiquitous Comput. [Internet]. 2012 [cited 2013 Dec 20];17:771–82. Available from: http://link.springer.com/10.1007/s00779-012-0626-y

53. Omkar SN, Vyas K, Vikranth HN. Time-frequency analysis of human motion during rhythmic exercises. Eng Med Bio Soc EMBC, 2011 Ann Int Conf IEEE. IEEE; 2011. p. 1279–82.

54. Arai T, Obuchi S, Shiba Y, Omuro K, Inaba Y, Kojima M. The validity of an assessment of maximum angular velocity of knee extension (KE) using a gyroscope. Arch Gerontol Geriatr. [Internet]. Elsevier; 2017;54:e175–80. Available from: http://dx.doi.org/10.1016/j.archger.2011.10.012

55. Chakraborty S, Jamthe A, Ghosh SK, Agrawal DP. Determining athlete's injury with wireless body area sensor network-based overhead squat testing. Mob Ad-Hoc Sens Syst (MASS), 2013 IEEE 10th Int Conf IEEE; 2013. p. 439–40.

56. Fitzgerald D, Foody J, Kelly D, Ward T, Markham C, McDonald J. Development of a wearable motion capture suit and virtual reality biofeedback system for the instruction and analysis of sports rehabilitation. Eng Med Biol Soc EMBS. 2007. p. 4870–4.

57. Ai QS, Chen L, Liu Q, Zou L. Rehabilitation assessment for lower limb disability based on multi-disciplinary approaches. Australas Phys Eng Sci Med. Springer; 2014;37:355–65.

58. Giggins O, Kelly D, Caulfield B. Evaluating rehabilitation exercise performance using a single inertial measurement unit. 7th Int Conf Pervasive Comput Technol Healthc Work. [Internet]. 2013 [cited 2014 Nov 13]. p. 49–56. Available from: http://eudl.eu/doi/10.4108/pervasivehealth.2013.252061

59. Chen K-H, Chen P-C, Liu K-C, Chan C-T. Wearable sensor-based rehabilitation exercise assessment for knee osteoarthritis. Sensors (Basel). [Internet]. Springer; 2015;15:4193–211. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4367405/

60. Houmanfar R, Karg M, Kulić D. Movement analysis of rehabilitation exercises: Distance metrics for measuring patient progress. IEEE Syst J. IEEE; 2016;10:1014–25.

61. Chen K-H, Tseng W-C, Liu K-C, Chan C-T. Using gyroscopes and accelerometers as a practical rehabilitation monitor system after total knee arthroplasty. 2015 IEEE MTT-S 2015 Int Microw Work Ser RF Wirel Technol Biomed Healthc Appl. IEEE; 2015. p. 58–9.

62. Streiner DL, Norman GR. Health Measurement Scales: A practical guide to their development and use. 5th ed. Oxford University Press. 2008.

63. Tuncel O, Altun K, Barshan B. Classifying human leg motions with uniaxial piezoelectric gyroscopes. Sensors (Basel). [Internet]. MDPI; 2009;9:8508–46. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3260598/

64. Chen P, Huang C-N, Chen I, Chan C. A rehabilitation exercise assessment system based on wearable sensors for knee osteoarthritis. Int Conf Smart Homes Heal Telemat. Springer; 2013. p. 267–72.

65. Ahmadi A, Mitchell E, Richter C, Destelle F, Gowing M, O'Connor NE. Automatic activity classification and movement assessment during a sports training session using wearable inertial sensors. IEEE Internet Things J. 2015;2:23–32.

66. Conger SA, Guo J, Fulkerson SM, Pedigo L, Chen H, Bassett DR. Objective assessment of strength training exercises using a wrist-worn accelerometer. Med Sci Sports Exerc. [Internet]. 2016;48:1847–55. Available from: http://europepmc.org/abstract/MED/27054678

67. O'Reilly M, Whelan D, Ward TE, Delahunt E, Caulfield B. Technology in S&C: Tracking lower limb exercises with wearable sensors. J Strength Cond Res. [Internet]. 2017;pre-print. Available from: http://insights.ovid.com/crossref?an=00124278-900000000-96098

68. Dominguez Veiga JJ, O'Reilly M, Whelan D, Caulfield B, Ward TE. Feature-free activity classification of inertial sensor data with machine vision techniques: Method, development, and evaluation. JMIR Mhealth Uhealth [Internet]. 2017;5:e115. Available from: http://mhealth.jmir.org/2017/8/e115/

69. Kianifar R, Lee A, Raina S, Kulić D. Classification of squat quality with inertial measurement units in the single leg squat mobility test. 38th Annu Int Conf IEEE Eng Med Biol Soc. 2016. p. 6273–6.

70. Whelan DF, O'Reilly MA, Ward TE, Delahunt E, Caulfield B. Technology in rehabilitation: Evaluating the single leg squat exercise with wearable inertial measurement units. Methods Inf Med. [Internet]. Schattauer Publishers; 2016;55:[Epub ahead of print]. Available from: http://dx.doi.org/10.3414/ME16-02-0002

71. O'Reilly MA, Whelan DF, Ward TE, Delahunt E, Caulfield B. Classification of lunge biomechanics with multiple and individual inertial measurement units. Sport Biomech. 2017;16:342–60.

72. O'Reilly MA, Whelan DF, Ward TE, Delahunt E, Caulfield BM. Technology in S&C: Assessing bodyweight squat technique with wearable sensors. J Strength Cond Res. 2017;

73. O'Reilly MA, Whelan DF, Ward TE, Delahunt E, Caulfield B. Technology in rehabilitation: Comparing personalised and global classification methodologies in evaluating the squat exercise with wearable imus. Methods Inf Med. 2017;56:1–9.

74. O'Reilly MA, Whelan DF, Ward TE, Delahunt E, Caulfield BM. Classification of deadlift biomechanics with wearable inertial measurement units. J Biomech. 2017;[ePub ahea:155–61.

75. O'Reilly M, Duffin J, Ward T, Caulfield B. Mobile app to streamline the development of wearable sensor-based exercise biofeedback systems: System development and evaluation. JMIR Rehabil Assist Technol [Internet]. 2017;4:e9. Available from: http://rehab.jmir.org/2017/2/e9/

76. Kotsiantis SB. Supervised machine learning: A review of classification techniques [Internet]. Informatica. 2007. Available from: http://books.google.com/books?hl=en&lr=&id=vLiTXDHr_sYC&oi=fnd&pg=PA3&dq=survey+machine+learning&ots=CVsyuwYHjo&sig=A6wYWvywU8XTc7Dzp8ZdKJaW7rc%5Cnpapers://5e3e5e59-48a2-47c1-b6b1-a778137d3ec1/Paper/p800%5Cnhttp://www.informatica.si/PDF/31-3/11_Kotsiantis

77. Zeng M, Nguyen LT, Yu B, Mengshoel OJ, Zhu J, Wu P. Convolutional neural networks for human activity recognition using mobile sensors. Pro 6th Int Conf Mob Comput Appl Serv. [Internet]. 2014. p. 197–205. Available from: https://www.computer.org/csdl/proceedings/mobicase/2014/024/00/07026300-abs.html

78. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R. Caffe: convolutional architecture for fast feature embedding. ACM Int Conf Multimed. 2014;675–8.

79. Whelan DF, O'Reilly MA, Huang B, Giggins OM, Kechadi T, Caulfield B. Leveraging IMU data for accurate exercise performance classification and musculoskeletal injury risk screening. 38th Annu Int Conf IEEE Eng Med Biol Soc 2016 [Internet]. IEEE; 2016. p. 1–6. Available from: http://dx.doi.org/10.3414/ME16-02-0002

80. Breiman L. Random forests. Mach Learn. [Internet]. 2001;45:5–32. Available from: http://link.springer.com/article/10.1023/A:1010933404324

81. Hearst MA, Dumais ST, Osman E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst. 1998;13:18–28.

82. Dudani SA. The Distance-Weighted k-Nearest-Neighbor Rule. IEEE Trans Syst Man Cybern. 1976;SMC-6:325–7.

Tables

Table 1: Literature search strategy.

Exercise	"exercise" OR "rehab*" OR "weight training" OR "motor activity" OR "personal" OR "strength" OR "conditioning" OR "hypertrophy" OR "gym" OR "weight lifting" OR "resistance" OR "training"
	AND
Lower body	"lower body" OR "lower extremity" OR "leg" OR "thigh" OR "shank" OR "ankle" OR "foot" OR "joint"
	AND
Movement monitoring	"monitor" OR "motion" OR "classif*" OR "recogn*" OR "evaluat*" OR "posture" OR "sensing" OR "assess*" OR "quantification" OR "biomech*" OR "tracking" OR "quality" OR "kinematics" OR "biofeedback"
	AND
Inertial measurement units	"inertial sensor" OR"gyroscop*" OR "IMU" OR "inertial measurement units" OR "wearable" OR "acceleromet*" OR " sensor system" OR "sensor network" OR "magnetometer" OR "MEMS" OR "smartphone" OR "mobile" OR "wireless"
	AND NOT
	"robot" OR "exoskeleton"

Table 2: Inclusion and exclusion criteria for studies.

Inclusion criteria	Exclusion criteria
The articles contain a system for exercise analysis using IMUs. The system is intended for monitoring repetition-based targeted exercises for the lower limb (e.g. squats, deadlifts, single leg squats, lunges, straight leg raises, and jumps), or analysing rehabilitation, workplace or strength and conditioning exercises The system included detection of exercises and/or quantification of exercise volume and/or analysis of exercise technique or performance measures. Articles were published in the last 10 years. Articles were written in the English language.	Systematic reviews and literature reviews. Books and other non-peer reviewed literature. Studies evaluating robotic systems or exoskeletons. Studies investigating human activity recognition in non-rehabilitative or strength and conditioning settings (i.e. in the 'real-world'). Studies evaluating pathological groups only. Sensing modality used was not a wearable accelerometer, gyroscope, magnetometer or combination of those (IMU). Study only concerns non-repetition based targeted exercises e.g. running, walking, gait, balance. Study concerns non-human, animal subjects. Study only evaluates 'user experience' with the system or the effect of the system's feedback on users.

IMU = inertial measurement unit

Table 3. Risk of bias assessment of the included studies based on the modified STROBE criteria [37].

Study	1	2	3	4	5	6	7	8	9	10	Quality
Ahmadi et al. 2014 [65]	1	1	0	1	1	0	0	0	0	1	Low
Ai et al. 2014 [57]	0	0	0	1	1	1	0	0	0	1	Low
Arai et al. 2012 [54]	1	0	1	1	1	1	0	0	1	1	Low
Bo et al. 2011 [51]	0	0	0	1	1	0	1	0	0	0	Low
Bolink et al. 2016 [39]	1	1	1	1	1	1	1	1	1	1	High
Bonnet et al. 2011 [7]	1	1	0	1	1	1	0	0	0	1	Low
Bonnet et al. 2013 [43]	1	1	1	1	1	1	1	1	1	1	High
Chakraborty et al. 2013 [55]	0	0	0	0	0	0	0	0	0	0	Low
Chang et al. 2007 [17]	1	0	0	0	1	0	1	0	0	0	Low
Charlton et al. 2015 [50]	1	1	1	1	1	1	1	1	1	1	High
Chen et al. 2013 [64]	1	0	0	1	1	0	1	0	0	0	Low
Chen et al. 2015 [61]	0	0	0	0	0	0	0	0	0	0	Low
Conger et al. 2016 [66]	1	1	1	1	1	1	1	1	1	1	High
Dominguez-Veiga et al. 2017 [68]	1	1	0	1	1	1	1	1	1	1	High
Faber et al. 2015 [40]	1	1	0	1	1	1	1	1	1	1	High
Fitzgerald et al. 2007 [56]	0	0	0	0	0	0	0	0	0	1	Low
Giggins et al. 2013 [18]	0	1	1	1	1	1	1	0	0	1	Low
Giggins et al. 2014 [3]	1	1	1	1	1	1	1	1	1	1	High
Giggins et al. 2014 [58]	1	1	1	1	1	1	1	1	1	0	High
Gleadhill et al. 2016 [21]	1	1	1	1	1	1	1	0	1	1	High
Gordon et al. 2012 [45]	0	1	0	1	1	0	1	0	0	0	Low
Haladjian et al. 2015 [38]	0	0	0	0	0	1	0	0	0	1	Low
Houmanfar et al. 2016 [60]	1	1	0	1	1	1	1	0	1	0	Low
Kianifir et al. 2016 [69]	1	1	1	1	1	1	1	0	0	0	Low
Lin and Kulic 2012 [20]	1	1	1	1	1	1	1	0	1	0	High
Mehta et al. 2016 [41]	1	1	1	1	1	1	1	1	1	0	High
Morales et al. 2017 [42]	1	1	1	1	1	1	1	1	1	1	High
Morris et al. 2014 [14]	1	1	0	1	1	1	1	1	0	0	Low
O’Reilly et al. 2017 [67]	1	1	0	1	1	1	1	1	1	1	High
O’Reilly et al. 2017 [71]	1	1	0	1	1	1	1	1	1	1	High
O’Reilly et al. 2017 [72]	1	1	0	1	1	1	1	1	1	1	High
O’Reilly et al. 2017 [73]	1	1	1	1	1	1	1	1	1	0	High
O’Reilly et al. 2017 [74]	0	1	0	1	1	1	1	1	1	1	High
O’Reilly et al. 2017 [75]	1	1	1	1	1	1	1	1	1	1	High
Omkar et al. 2011 [53]	1	1	0	1	1	1	1	0	0	1	Low
Papi et al. 2015 [48]	1	1	0	1	1	1	1	1	1	1	High
Patterson and Caulfield 2010 [24]	1	1	0	1	1	1	1	1	0	1	High
Pernek et al. 2012 [52]	1	1	0	1	1	1	1	0	0	1	Low
Quaglierella et al. 2010 [44]	1	1	1	1	1	1	1	1	0	1	High
Rawson and Walsh 2010 [47]	1	1	1	1	1	1	1	0	0	1	High
Setuain et al. 2015 [46]	1	1	1	1	1	1	1	0	1	0	High
Setuain et al. 2015 [22]	1	1	0	1	1	1	1	1	1	1	High
Taylor et al. 2010 [23]	0	1	0	1	1	0	1	1	0	0	Low
Tunçel et al. 2009 [63]	0	1	0	1	1	0	1	0	1	1	Low
Whelan et al. 2016 [70]	1	1	0	1	1	1	1	1	1	1	High
Yurtman and Barshan 2014 [19]	1	1	0	1	1	1	0	1	1	0	Low
Zijlstra et al. 2010 [49]	1	1	1	1	1	1	1	0	1	0	High

Items legend: 1. Provide in the abstract an informative and balanced summary of what was done and what was found. 2. State specific objectives, including any prespecified hypotheses. 3. Give the eligibility criteria, and the sources and methods of selection of participants. 4. For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group. 5. Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why. 6. Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders. 7. Summarise key results with reference to study objectives. 8. Discuss limitations of the study, considering sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias. 9. Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence. 10. Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based. STROBE = STrengthening the Reporting of OBservational studies in Epidemiology.

Table 4: Sensing set-ups evaluated.

Sensor set-up	**Studies**
Multiple sensor units (n=24)	[17],[19],[63],[65],[44],[20],[45],[51],[55],[64],[61],[23],[40],[48],[38],[69],[60]
Single sensor units (n=19)	[14],[52],[58],[24],[53],[7],[43],[57],[39],[41],[46],[54],[42],[50],[22],[75],[68]
Comparison of multiple and single sensor units (n=12)	[18],[47],[3],[49],[21],[40],[67],[70],[71],[72],[73],[74]

Table 5: Sensors used in each study included in this review.

Study	Accelerometer	Gyroscope	Magnetometer	Other
Ahmadi et al. 2014 [65]	■	■		
Ai et al. 2014 [57]	■	■	■	
Arai et al. 2012 [54]		■		
Bo et al. 2011 [51]	■	■		■
Bolink et al. 2016 [39]	■	■	■	
Bonnet et al. 2011 [7]	■	■		
Bonnet et al. 2013 [43]	■	■		
Chakraborty et al. 2013 [55]	■	■		
Chang et al. 2007 [17]	■			
Charlton et al. 2015 [50]	■	■	■	
Chen et al. 2013 [64]	■			
Chen et al. 2015 [61]	■	■		
Conger et al. 2016 [66]	■			
Dominguez-Veiga et al. 2017 [68]	■	■		
Faber et al. 2015 [40]	■	■		
Fitzgerald et al. 2007 [56]	■	■	■	
Giggins et al. 2013 [18]	■	■		
Giggins et al. 2014 [3]	■	■		
Giggins et al. 2014 [58]	■	■		
Gleadhill et al. 2016 [21]	■	■		
Gordon et al. 2012 [45]	■	■		
Haladjian et al. 2015 [38]	■	■		
Houmanfar et al. 2016 [60]	■	■	■	
Kianifir et al. 2016 [69]	■	■	■	
Lin and Kulic 2012 [20]	■	■		
Mehta et al. 2016 [41]	■	■		
Morales et al. 2017 [42]	■	■	■	
Morris et al. 2014 [14]	■	■		
O'Reilly et al. 2017 [67]	■	■	■	
O'Reilly et al. 2017 [71]	■	■	■	
O'Reilly et al. 2017 [72]	■	■	■	
O'Reilly et al. 2017 [73]	■	■	■	
O'Reilly et al. 2017 [74]	■	■	■	
O'Reilly et al. 2017 [75]	■	■	■	
Omkar et al. 2011 [53]	■	■		
Papi et al. 2015 [48]	■	■		■
Patterson and Caulfield 2010 [24]	■			
Pernek et al. 2012 [52]	■			
Quaglierella et al. 2010 [44]	■			
Rawson and Walsh 2010 [47]	■			
Setuain et al. 2015 [46]	■	■	■	
Setuain et al. 2015 [22]	■	■	■	
Taylor et al. 2010 [23]	■			
Tunçel et al. 2009 [63]		■		
Whelan et al. 2016 [70]	■	■	■	
Yurtman and Barshan 2014 [19]	■	■	■	
Zijlstra et al. 2010 [49]	■	■	■	

Table 6: Exercises which have been investigated during the studies included in this review and the study type which they were included in.

	Exercise	Measurement validation	Exercise detection	Movement classification
SJ-UP	Lying hip abduction	[58], [50]	[18], [19]	[3], [19]
	Lying hip extension	[58], [50]	[18], [19]	[3], [19]
	Lying knee flexion (supine)	[41]		
	Inner range quads	[58]	[18], [59]	[3], [59]
	Seated knee extension	[58], [52], [54]	[18], [63],	[3]
	Seated knee flexion	[20]		
	Lying straight leg raise	[58], [20]	[18], [19], [59]	[3], [19],[59], [23]
	Standing calf raise	[52]	[17], [66]	
	Seated straight leg raise	[60]	[19],	[19]
	Standing straight leg raise	[20]	[63]	
	Standing knee flexion/extension	[38]	[63]	[23]
	Standing hip extension		[63]	
	Standing hip abduction		[63]	[23]
	Standing leg curl	[52]		
	Seated calf raises	[52]		
	Lying leg curl	[52], [47]		
	Seated resisted knee extension	[52], [47]		
	Ankle dorsi/plantarflexion	[57]		
	Ankle internal/external rotation	[57]		
	Ankle inversion/eversion	[57]		
	Seated hip internal/external rotation	[50]		
	Supine hip internal/external rotation	[50]		
SJ-MP	Lying straight diagonal leg raise	[20]		
	Standing circle trace (hip)	[20]		
	Lying circle trace (hip)	[20]		
MJ-UP	Heel slides	[58], [61]	[18]	[3]
	Lying hip & knee flexion	[58], [20], [60]	[18]	[3],
BMJ-MP	Sit to stand	[49], [20], [51], [48], [39]		
	Leg press	[52]		
	Lunge	[56], [42]	[66], [67], [68]	[71], [75]
	Kicking		[65]	
	Deadlift	[21]	[17]	[75]

	Mini-squats		[59]	[59]
	Squats	[52], [47], [7], [43], [20], [51], [38], [60]	[14], [63], [66], [67], [68]	[72], [73], [75]
	Barbell deadlifts		[67], [68]	[74], [75]
	Overhead squats	[55]		
	Kettlebell swing		[14]	
	Sun salutation	[53]		
	Hang clean	[45]		
	Block step up	[39]		
	Single leg squats		[67], [68]	[69], [70]
	Box lift	[40]		
	Stoop box lift	[40]		
	Squat box lift	[40]		
	One leg hops	[38]		
	Side hops	[38]		
	Box jump		[65]	
	Bilateral squat jumps	[44]		
	Bilateral Countermovement Jumps	[45], [44]		
	Drop jumps	[24], [46], [22],		
	Unilateral drop jump	[44], [22]		
	Unilateral countermovement jumps	[44], [22]		
	Tuck jumps		[67],[68]	

*Key: SJ-UP = Single-joint, uni-planar, SJ-MP = Single-joint, multi-plane, MJ-UP = Multi-joint, uni-planar, MJ-MP = Multi-joint, multi-plane

Table 7: Summary of studies assessing concurrent validity of wearable sensor based system to standard clinical measure or biomechanical gold standard.

Study	Sensor set-up and placement	Sample	Outcomes	Gold standard/ comparator	Findings
Lin and Kulic 2012 [20].	3 x tri-axial accel + gyro (trunk, thigh, shank)	20 (8 females, 12 females), injury-free	Joint angles during 9 lower limb rehabilitation exercises	Optoelectronic motion capture	The average RMSE between a motion capture system and their wearable system was 6.5 º across all exercises.
Haladjian et al. 2015 [38].	2 x tri-axial accel + gyro (thigh & shank)	4 (2 females, 2 males), following ACL surgery	Sagittal plane knee joint angle	Goniometer	There was a 5 º difference in agreement between the 'KneeHapp' system and the goniometer.
Bolink et al. 2016 [39].	1 x tri-axial accel + gyro + mag (lumbar)	17 (8 females, 9 males), injury-free	Pelvic orientation angles during the sit to stand and block step up exercises	Optoelectronic motion capture	Frontal plane pelvic angle estimations achieved a RMSE in the range of 2.7° to 4.5° and sagittal plane measurements achieved a RMSE in the range of 2.7° to 8.9° when compared with optoelectronic motion capture
Faber et al. 2015 [40].	1 x accel + gyro (multiple locations from C7-MPSIS)	20 (10 females, 10 males), injury-free	Which location optimally agreed with an optoelectronic motion capture system's calculation for trunk inclination? The data used was from a variety of box lifting exercises.	Optoelectronic motion capture	They concluded that regardless of participant's sex or lifting style, the optimal sensing unit location for the measurement of trunk inclination is at about 25% of the distance from the sacrum to C7.
Mehta et al.	1 x iPhone (tri-axial accel +	60 (sex not reported),	Knee flexion and	Standard goniometry	They showed the mobile application

2016 [41].	gyro)	following total knee replacement or with knee osteoarthiritis	extension ROM		allowed for a smaller minimal detectable change than goniometry.
Morales et al. 2017 [42].	1 x smartphone (accel + gyro + mag)	33 (sex not reported), injury-free	Inclination of the tibia during the weight bearing lunge exercise	Tape measure test, goniometry and the leg motion system	They found no significant differences between any of the measurement techniques.
Bonnet et al. 2011, 2013 [7,43].	1 x tri-axial accel + gyro (lumbar)	10 (4 females, 6 males), injury free	Sagittal hip, knee and joint angles during a bodyweight squat exercise	Optoelectronic motion capture	Their most recent predictive algorithm had a root mean square difference of 3.2 º,· 2 º and 3.1 º for ankle, knee and hip angles respectively when compared with both a robot model and with 8 healthy human participants [43].
Patterson and Caulfield 2010 [24].	1 x tri-axial accel (ankle)	20 (14 females, 6 males), injury free	Reactive strength index during the drop jump exercise	Force plate data	Pearson's product correlation of 0.9816 in computing reactive strength index during the drop jump exercise
Quaglieralla et al. 2010 [44].	2 x tri-axial accel (left and right ankle)	51, (26 injury-free, 25 following surgery for Achilles tendon rupture; 51 males)	Flight time during countermovement jumps and squat jumps	Force plate	Spearman's coefficient was found to be greater than 0.95 in this case.
Gordon et al. 2012 [45].	2 x tri-axial accel + gyro (trunk and barbell)	1 (male), injury-free	Mean percentage error for the following measurements: Peak velocity, time to peak velocity, peak power,	Optoelectronic motion capture system and force plates	The temporal measures had the lowest mean percentage error with time to peak velocity and time to peak power having an error of just 0.034% and 1.01% respectively. The kinetic measures had a

			time to peak power and force at peak power.		larger error with peak power and peak force both resulting in a 12.5% error versus the force plates and motion capture system.
Setuain et al. 2015 [46].	1 x tri-axial accel + gyro + mag (lumbar)	17 (8 females, 9 males), injury-free	Vertical force derived from IMU	Force plates	Several biomechanical variables such as the resultant force–time curve patterns in drop jumps, unilateral drop jumps and unilateral countermovement jumps can be reliably measured with a lumbar worn IMU
Rawson and Walsh 2010 [47].	3 x uni-axial accel (wrist, waist and ankle)	30 (15 females, 15 males), injury free	Activity counts during the squat, leg extension and leg curl exercises	Cosmed™ system (COSMED, Rome, Italy)	Activity counts were correlated with energy expenditure as computed by a cosmed™ system (COSMED, Rome, Italy). Thirty healthy participants were recruited and a primary finding of the study was that a regression equation which inputs included sex, fat-free mass, and counts of activity from the waist accelerometer explained 90% (R^2 = 0.90) of the variance in energy expenditure as measured by the cosmed™ system.
Papi et al. 2015 [48].	1x tri-axial accel + gyro (waist), 1x tri-axial accel (waist) + bend sensor also used (knee)	14 (7 females, 7 males), injury-free	Total time taken to complete a five time sit to stand test	Optoelectronic motion capture	The waist worn sensing unit was found to have a 0.86 RMSE versus the measure from a motion capture system
Ziljstra et al.	3 x tri-axial accel + gyro + mag	17 (10 females, 7	Vertical power during	Optoelectronic motion	They used Pearson’s correlation to

2010 [49].	(sternum, pelvis, SIPS)	males), injury free	the sit to stand test	capture + force plates	compare each sensor position's power output to force plate data and found an R^2 of 0.984 at the body's estimated centre of mass.
Gleadhill et al. 2016 [21].	3 x tri-axial accel+gyro on spine (C7, T12 and S1)	11 (1 female, 10 males), injury free	Temporal features from accelerometers during fifteen variations of the deadlift exercise.	Optoelectronic motion capture	The average Pearson's correlation with a motion capture system was R^2 = 0.9997 for sagittal plane accelerometer peaks.
Charlton et al. 2015 [50].	1 x smartphone (accel + gyro + mag)	20 (males), injury free	Hip ROM (flexion, abduction, adduction, supine internal and external rotation and sitting internal and external rotation)	Optoelectronic motion capture	The Smartphone demonstrated good to excellent reliability (ICCs > 0.75) for four out of the seven movements, and moderate to good reliability for the remaining three movements (ICC = 0.63–0.68)

Accel = accelerometer, ACL = anterior cruciate ligament, C7 = Cervical vertebrae 7, gyro = gyroscope, ICC = intra-class correlation, IMU = inertial measurement unit, mag = magnetometer, MPSIS = midpoint between the posterior superior iliac spines, SIPS = posterior superior iliac spine, RMSE = root mean square error, ROM = range of motion.

Table 8: Summary of studies assessing construct validity of wearable sensor based systems.

Study	Sensor set-up and placement	Sample	Outcomes	Construct validity type	Comparator	Findings
Bo et al. 2011 [51]	2 x tri-axial accel 2 x dual-axial gyro + Microsoft Kinect (thigh and shank)	Not described	Knee angles during sit to stand and squat.	Convergent	Microsoft Kinect	High potential for fusion of Kinect and inertial sensors for more accurate joint angle measurement
Pernek et al. 2012 [52]	1 x smartphone w/ tri-axial accel (on weights stack or ankle)	10 (4 females, 6 males), injury-free	Detection of exercise repetitions and the start/end of repetitions.	Convergent	Manual extraction of repetitions by authors.	99% accuracy in repetition detection, 89% accuracy in detecting start and end points of repetitions.
Omkar et al. 2011 [53]	1 x tri-axial accel + gyro (lumbar)	11 (4 females, 7 males), injury-free	Grace and consistency during rhythmic exercise	Convergent	Visual analysis of each participant's sequence by yoga expert	Found performance of 2 participants to be significantly worse than the others (more jerks and halts).
Arai et al. 2012 [54]	1 x tri-axial gyro (shank)	105 (55 females, 50 males), elderly, injury-free	Physical function and self-efficacy	Convergent	Functional performance measurements, a self-efficacy scale and HRQOL.	Gyroscope peaks correlated with some physical functions such as muscle strength ($r = 0.304$, $p < 0.01$), and walking velocity ($r = 0.543$, $p < 0.001$). In addition, the joint angular velocity was significantly correlated with self-efficacy ($r = 0.219–0.329$, $p < 0.01–0.05$) and HRQOL ($r = 0.207–0.359$, $p < 0.01–0.05$).
Chakraborty et al. 2013 [55]	Xsens™ moCap suit	6 (sex not reported), undergoing rehab of lower limb	Body posture during overhead squat task	Known-groups	Individual's measures pre & post injury	Not described.

Fitzgerald et al. 2007 [56]	Xsens™ moCap suit	2 (sex not reported), one injury-free, one 15 weeks post MCL tear	Body posture during straight line lunge	Known-groups	Comparison of injured and uninjured individual	Greater range of trunk flexion/extension, thigh internal/external rotation and trunk flexion/extension for injured athlete.
Ai et al. 2014 [57]	1 x tri-axial accel + gyro + mag (instep of foot OR shank)	3 (males;1 healthy, 1 polymyotosis, 1 chronic lower back pain)	ROM, movement smoothness, trajectory error	Known-groups	Comparison of each group's results	Proof of concept for tracking ankle exercises with IMUs shown via each participant's differing trajectories.
Giggins et al. 2014 [58]	1 x tri-axial accel + gyro (shin)	9 (5 females, 4 males), injury-free	Signal features when exercises completed with correct technique and aberrant technique.	Known-groups	Comparison of features when exercises completed with acceptable and aberrant technique.	A number of significantly different features found across all exercises and all deviations/known groups.
Setuain et al. 2015 [22]	1 x accel + gyro + mag (lumbar)	22, (sex not reported; 6 ACL reconstructed and 16 injury-free)	Signal features during a battery of vertical jumping tests.	Known-groups.	Comparison of features when exercises completed by ACL reconstructed and injury-free group.	The ACL-reconstructed male athletes did not show any significant ($P < .05$) residual jumping biomechanical deficits regarding the measured variables compared to players who had not suffered this knee injury. A dominance effect was observed among non-ACL reconstructed controls but not among their ACL-reconstructed counterparts ($P < .05$).
Chen et al. 2015 [59]	2 x tri-axial accel + gyro (thigh and shank)	10, (5 females, 5 males; 5 injury-free + 5,	Knee ROM during heel slides tested 1 day pre, 1 day post,	Longitudinal	Individual's known	Knee ROM return to baseline after 6 weeks effectively measured with inertial sensors

		rehab after total knee arthroplasty)	2 weeks post and 6 weeks post total knee arthroplasty.		improvement in ROM during rehab.	
Houmanfar et al. 2016 [60]	2 x tri-axial accel + gyro + mag (thigh and shank)	28, (sex not reported; 18, rehab following knee/hip replacement, 10, injury-free)	Difference between patient's data (joint angle, velocity, acceleration) and healthy norms throughout rehabilitation.	Longitudinal	Individual's known improvement in ROM during rehabilitation.	The results show that the IMU measures are able to capture the trend of patient improvement over the course of rehabilitation.

Accel = accelerometer, ACL = anterior cruciate ligament, gyro = gyroscope, HRQOL = Health related quality of life, IMU = inertial measurement unit, mag = magnetometer, rehab = rehabilitation, ROM = range of motion.

Table 9: Summary of wearable inertial sensor exercise detection systems for lower limb exercises.

Study	Sensing set-up	Participants	Relevant exercises	Methods: classification	Methods: cross-validation	Results:
Chen et al. 2013 [64]	3 x accel, (trunk, thigh, shank)	10 (5 females, 5 males), injury-free, 10 reps per exercise	Inner range quads Straight leg raise Quadriceps mini-squats	Feature based classification. Decision tree.	10-CV	10-CV acc = 99.29%
Chang et al. 2007 [17]	2 x accel, (right hand glove and hip worn posture clip)	10 (2 females, 8 males), healthy, 15 reps per exercise	Deadlift Standing calf raise	Feature based classification: Hidden Markov models, Naïve bayes.	User specific CV and LOSOCV	User specific: acc = 95% LOSOCV acc=85%
Giggins et al. 2014 [18]	3 x tri-axial accel + gyro (foot, shank and thigh)	58 (39 females, 19 males), rehab, 10 reps per exercise	Heel slides Hip abduction Hip extension Hip flexion Inner range quads Knee extension Straight leg raise	Feature based classification. Logistical regression.	LOSOCV Multi and single sensor set-ups.	LOSOCV all 3 sensors: acc = 94% shank sensor: acc = 95%
Yurtman and Barshan. 2014 [19]	5 x tri-axial accel + gyro + mag (trunk, thighs, shanks OR thigh, upper arm, lower arm, trunk, shoulder)	5 (2 females, 3 males), rehab, 30 reps per exercise	Lying leg raise Lying hip abduction Lying hip extension Seated straight leg raise	Dynamic time warping.	LOEOCV	LOEOCV: acc = 93%
Morris et al. 2014. [14]	1 x tri-axial accel + gyro (wrist)	20 (8 females, 12 males), healthy, 20 reps per exercise	Jumping jack Kettlebell Swing Squat	Feature based classification. Support vector machine.	LOOCV	LOOCV: acc = 96%
Tuncel et al. 2009 [63]	2 x uni-axial gyro (right thigh and shank)	1 (male), healthy, 8 repetitions per exercise	Standing knee flexion Standing flexed leg raise Standing straight leg raise Standing R & L straight leg	Feature based classification. Support vector machine Dynamic time warping	RRSS P-fold -CV and LOOCV	RRSS: BDM acc: 98%, P-fold-CV: BDM acc: 99.1%, LOOCV: BDM acc: 99.1%

			Hip extension Standing hip abduction Bodyweight exercises Seated knee extension	Artificial neural network Radial basis function Bayesian decision making Least squares method K-nearest-neighbours		
Ahmadi et al. 2015 [65]	2 x tri-axial gyro (right thigh and shank)	10 (sex not reported), 9 healthy, 1 injured, 30 s per exercise	Box jump Kicking	Feature based classification Random forests	10-CV	Average F1 score of 97% in detecting each exercise.
Conger et al. 2016 [66]	1 x tri-axial accel (non-dominant wrist)	60 (sex not reported), healthy, 10 reps per exercise	Squats Lunges Calf Raises	Cosine similarity and feature based classification Support vector machine	LOSOCV	SVM method detected the lower-limb exercises with 81% acc. The cosine similarity method produced 85% accuracy.
O'Reilly et al. 2016 [67]	5 x tri-axial accel + gyro + mag (lumbar, thigh L&R, shank L&R)	82 (23 females, 59 males) healthy, 10 reps per exercise	Squats Lunges Single leg squats Deadlifts Tuck Jumps	Feature based classification. Random forests.	LOSOCV Multi and single sensor set-ups.	The exercises were detected with 99% acc when using signals from all five IMUs, 98% when using signals from the thigh and lumbar IMUs and 98% with just a single IMU on the shank.
Dominguez-Veiga et al. 2017 [68]	5 x tri-axial accel + gyro + mag (lumbar, thigh L&R, shank L&R)	82 (23 females, 59 males), healthy, 10 reps per exercise	Squats Lunges Single leg squats Deadlifts Tuck Jumps	Feature free classification. Convolutional neural network	Batch output training-CV	acc: 95.89%

Acc = accuracy, accel = accelerometer, ACL = anterior cruciate ligament, gyro = gyroscope, IMU = inertial measurement unit, mag = magnetometer, rehab = rehabilitation, ROM = range of motion, L = left, R = right, reps = repetitions, CV = cross-validation, LOSOCV = leave-one-subject-out-cross-validation, LOOCV = leave-one-out-cross-validation, LOEOCV = leave-one-exercise-out-cross-validation, RRSS = repeated random sub-sampling, SVM = support vector machine

Table 10: Summary of wearable inertial sensor movement classification systems for lower limb exercises.

Study	**Sensing set-up**	**Participants**	**Relevant exercises**	**Performance classified**	**Methods: classification**	**Methods: Cross-validation**	**Results:**
Taylor et al. 2010 [23]	5 x accel, (trunk, thigh L&R, shank L&R)	9 (5 females, 4 males), injury-free, 60 reps per exercise	Standing hamstring curls Reverse hip abduction Straight leg raise	Multi-label technique classification	Global feature based classification. Adaboost.	LOSOCV	LOSOCV :mean sens for each deviation = 44.9%, mean spec for each deviation = 64.5%
Chen et al. 2013 [64]	3 x accel, (trunk, thigh and shank)	10 (5 females, 5 males), injury-free, 10 reps per exercise	Inner range quads Straight leg raise Quadriceps mini-squats	Binary technique classification	Global feature based classification. Decision tree.	10-CV	10-CV acc = 90.14%
Yurtman and Barshan. 2014 [19]	5 x IMUs, (trunk, thigh L&R, shank L&R)	5 (2 females, 3 males), rehab, 30 reps per exercise	Seated straight leg raise Lying leg raise Lying hip abduction Lying hip extension	Speed of repetition ROM	Dynamic time warping	LOECV	LOECV acc = 89%
Giggins et al. 2014 [3]	3 x accel + gyro, (thigh, shank, foot)	58 (39 females, 19 males), rehab, 10 reps per exercise	Heel slides Hip abduction Hip extension, Hip flexion Inner range quads Knee extension Straight leg raise	Binary technique classification Multi-label technique classification	Global feature based classification. Logistic regression. Multi and single sensor set-ups.	LOSOCV	Binary LOSOCV all 3 sensors: (acc = 81%, sens=79%, specificity = 70%) LOSOCV thigh sensor: (acc=82%, sens=72%, spec=83%)
Kianifir et al. 2016 [69]	3 x IMU, (trunk thigh, shank)	7 (1 females, 6 males), injury-free, 5 reps per exercise	Single leg squat	Binary technique classification Multi-label technique classification	Global feature based classification. Support vector machine, decision tree and logistic regression.	LOSOCV 10-CV	Binary LOSOCV: SVM acc=98.6%, Multi-label LOSOCV: Decision tree acc=73%
Whelan et al. 2016 [70]	3 x tri-axial accel + gyro + mag (lumbar, thigh L&R,	83 (23 females 60 males), injury-free, 10 reps	Single leg squat	Binary technique classification Detection of 6 deviations	Global feature based classification. Random Forests.	RRSS	3IMU system, binary RRSS: acc = 77%, 1 IMU system (Shank L), binary RRSS: acc =

	shank L&R)			(parallel binary classifiers)	Multi and single sensor set-ups.		76%. Deviation detection, 1 IMU system (Shank L), binary RRSS: acc = 66-75%.
O'Reilly et al. 2017. [71]	5 x tri-axial accel + gyro + mag (lumbar, thigh L&R, shank L&R)	80, (23 females, 57 males), injury-free, 10 reps of 'acceptable form' + 3 reps per exercise deviation	Lunges	Binary technique classification Multi-label technique classification	Global feature based classification. Random forests. Multi and single sensor set-ups.	LOSOCV	A single IMU system achieved 83% acc, 62% sens and 90% spec in binary classification and a five IMU system achieved 90% acc, 80% sens and 92% spec. A five IMU set-up can also detect specific deviations with 70% accuracy.
O'Reilly et al. 2017. [72]	5 x tri-axial accel + gyro + mag(lumbar, thigh L&R, shank L&R)	77, (22 females, 55 males), injury-free, 10 reps of 'acceptable form' + 3 reps per exercise deviation	Bodyweight squats	Binary technique classification Multi-label technique classification	Global feature based classification. Random forests. Multi and single sensor set-ups.	LOSOCV	Acceptable or aberrant BW squat technique can be detected with 98% accu, 96% sens and 99% spec when using features derived from all 5 IMUs. Detecting exact deviations from acceptable BW squatting technique can be achieved with 80% acc using a 5 IMU system and 72% acc when using a single IMU positioned on the R shank.
O'Reilly et al. 2017. [73]	5 x tri-axial accel + gyro + mag Lumbar Thigh L & R Shank L&R	55, (18 females, 37 males), injury-free, reps form a full 3RM test.	Barbell back squats	Binary technique classification Multi-label technique classification	Global and personalised feature based classification. Random forests. Multi and single sensor set-ups.	Global: LOSOCV Personalised: LOOCV	Global classification techniques produced poor acc, sens and spec scores in binary classification even with a 5 IMU set-up in both binary (acc: 64%, sens:

							70%, spec: 28%) and multi-class classification (acc: 59%, sens: 24%, spec: 84%). Personalised classification techniques 1 IMU system (L thigh) classification scores (acc: 81%, sens: 81%, spec: 84%) and multi-class scores (acc: 69%, sens: 70%, spec: 89%).
O'Reilly et al. 2017. [74]	5 x tri-axial accel + gyro + mag Lumbar Thigh L & R Shank L&R	55, (18 females, 37 males), injury-free, reps form a full 3RM test & 10 reps of 'acceptable form' + 3 reps per exercise deviation	Barbell deadlifts	Binary technique classification Multi-label technique classification	Global and personalised feature based classification. Random Forests. Multi and single sensor set-ups.	Global: LOSOCV Personalised: LOOCV	Binary classification with real deviations using personalised model (LOOCV): 5 IMU – acc: 84%, lumbar IMU – acc: 80%. Detection of exact real deviations using personalised model (LOOCV): 5 IMU – acc: 78%, lumbar IMU – acc: 75%.
O'Reilly et al. 2017. [75]	1 x tri-axial accel + gyro + mag Thigh L	15, (3 females, 12 males), injury-free, 40 reps per exercise.	Squats Lunges Single leg squats Deadlifts	Binary technique classification	Personalised feature based classification. Random forests.	Real world system evaluation	The personalised systems achieved 89.50% acc, with 90.00% sens and 89.00% spec

Acc = accuracy, accel = accelerometer, ACL = anterior cruciate ligament, gyro = gyroscope, IMU = inertial measurement unit, mag = magnetometer, rehab = rehabilitation, ROM = range of motion, L = left, R = right, reps = repetitions, CV = cross-validation, LOSOCV = leave-one-subject-out-cross-validation, LOOCV = leave-one-out-cross-validation, LOEOCV = leave-one-exercise-out-cross-validation, RRSS = repeated random sub-sampling, SVM = support vector machine, sens = sensitivity, spec = specificity, BW = bodyweight

Table 11: Summary of recommendations for future studies.

Area of consideration	Recommendations
Sensor set-up	Study design will benefit from collecting data with comprehensive set-ups (multiple IMUs at a variety of relevant anatomical locations). Analysis can then compare system efficacy when employing all data from the comprehensive set-up and a variety of subsets of these data. Engage target system users in advance of data collection to develop an understanding of their preferences regarding factors relating to sensor set-up e.g. cost, accuracy, usability and functionality.
Measurement validation	Studies assessing known groups validity e.g. to assess IMU systems' capacity to differentiate injured and non-injured individuals should recruit larger samples to allow for formal statistical analyses. There exists a need for predictive algorithms from IMU data to estimate kinetic exercise parameters and to assess their concurrent validity with force plate and optoelectronic motion capture data. An under investigated field is using a single IMU set-up to predict lower limb joint angles.
Exercise detection	Exercise detection systems' accuracies may be increased by collecting larger data sets from a greater number of participants. When assessing system efficacy via cross-validation techniques data from the same participant should not be included in both the training and test sets. Deep learning techniques such as convolutional neural networks and long-short-term memory networks may improve exercise detection efficacy.
Movement classification	Larger data sets, collected from more participants and inclusive of more exercise types and their associated technique deviations are necessary to further develop this area. If developing a global classification system, when assessing system efficacy via cross-validation techniques data from the same participant should not be included in both the training and test sets. Experimentally evaluating a system's real-world accuracy would also strengthen the literature. Deep learning techniques such as convolutional neural networks and long-short-term memory networks may improve movement classification efficacy. Personalised classification systems accuracy and efficiency outperforms global classification systems but require time and expertise to develop. Tools which streamline this development process should be investigated.

IMU = inertial measurement unit