Journal of Sports Sciences

Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance

--Manuscript Draft--

Full Title:	Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance
Manuscript Number:	RJSP-2018-0332R2
Article Type:	Original Manuscript
Keywords:	Sport movement classification; inertial sensors; computer vision; Machine learning; performance analysis.
Abstract:	Objective assessment of an athlete's performance is of importance in elite sports to facilitate detailed analysis. The implementation of automated detection and recognition of sport-specific movements overcomes the limitations associated with manual performance analysis methods. The object of this study was to systematically review the literature on machine and deep learning for sport-specific movement recognition using inertial measurement unit (IMU) and, or computer vision data inputs. A search of multiple databases was undertaken. Included studies must have investigated a sport-specific movement and analysed via machine or deep learning methods for model development. A total of 52 studies met the inclusion and exclusion criteria. Data pre-processing, processing, model development and evaluation methods varied across the studies. Model development for movement recognition were predominantly undertaken using supervised classification approaches. A kernel form of the Support Vector Machine algorithm was used in 53% of IMU and 50% of vision-based studies. Twelve studies used a deep learning method as a form of Convolutional Neural Network algorithm and one study also adopted a Long Short Term Memory architecture in their model. The adaptation of experimental set-up, data pre-processing, and model development methods are best considered in relation to the characteristics of the targeted sports movement(s).
Order of Authors:	Emily Elizabeth Cust
	Alice J Sweeting
	Kevin Ball
	Sam Robertson
Response to Reviewers:	The authorship team have read and responded to the comments of reviewer #3. The red coloured text in the revised manuscript highlights the new alterations and additions. Reviewer #1: The authors replied to my previous comments in a satisfactory way, then, I would approve the publication of this systematic review. Author's response: The authorship team thank Reviewer #1 for their previous constructive comments. Reviewer #3: I think two important datasets are missing here. oThe Volleyball dataset proposed by [1]. This dataset is for group activity recognition in sport footage. I think most of the team sport datasets contains multiple people, so group activity recognition is an important task in the team sport analysis. oNCAA Basketball dataset, this is a multi-person action video dataset in team sport context. [5] Author's response: We thank the reviewer for alerting us to these two papers. Given that they meet the requirements for inclusion, both these articles have now been included in the review. Tables 4, 7, 8 have been amended to include the relevant information. Also, these articles have been cited in the discussion section on lines 543 - 545. The Prisma flow diagram (Figure 1) has been updated and the study result numbers throughout this review have also been updated to reflect the additional articles. Reviewer #3:

One resource is missed here, MIT SLOAN SPORTS ANALYTICS Conference [2] is a one important source for recent works on sport analytics.
Author’s response:
The papers mentioned by the reviewer did not meet the whole inclusion and exclusion criteria for this review paper.

Reviewer #3:
Table 2 shows the inclusion and exclusion criteria for the search. In the Exclusion criteria, it has been mentioned that works with this condition are excluded:
"Solely investigated player field positional tracking methods using data such as X, Y coordinates or displacement without any form of sport-specific skill detection and classification associated to it" and "Used ball trajectory and audio cue data as the major determinant for event detection".
I don't understand why these works are excludes. I think that trajectories (Players X,Y coordinates) are a valuable source for activity recognition.[3][4]
Author’s response:
The papers mentioned by the reviewer did not meet the whole inclusion and exclusion criteria for this review paper.

Reviewer #3:
Missing reference: [6]
Author’s response:
This article has now been included in the review. Tables 4, 7, 8 have been amended to include the relevant information. Also, this article has been cited in the discussion section on lines 543 -545. The Prisma flow diagram (Figure 1) has been updated and the study result numbers throughout this review have also been updated to reflect the additional article.

Reviewer #3 references provided:
[1] Mostafa S. Ibrahim, Srikanth Muralidharan, Zhiwei Deng, Arash Vahdat, Greg Mori. A Hierarchical Deep Temporal Model for Group Activity Recognition. CVPR 2016.
[2] www.sloansportsconference.com
[3] N Mehrasa, Y Zhong, F Tung, L Bornn, G Mori. Deep Learning of Player Trajectory Representations for Team Activity Analysis. SLOAN 2018.
[4] Kuan-Chieh Wang and Richard Zemel. Classifying nba offensive plays using neural networks. In MIT SLOAN Sports Analytics Conference, 2016.
[5] Vignesh Ramanathan, Jonathan Huang, Sami Abu-El-Haija, Alexander Gorban, Kevin Murphy, and Li Fei-Fei. Detecting events and key actors in multi-person videos. CVPR 2016.
[6] Moumita Roy Tora, Jianhui Chen, James J. Little.Classification of Puck Possession Events in Ice Hockey. CVPR Workshop. 2017

Manuscript - with author details

Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance

Emily E. Cust[1, 2*], *Alice J. Sweeting*[1, 2], *Kevin Ball*[1] *and Sam Robertson*[1, 2]

Author details:

[1] Institute for Health and Sport (IHES), Victoria University, Melbourne, Australia

[2] Western Bulldogs Football Club, Footscray, Melbourne, Australia

Author ORCID

Sam Robertson 0000-0002-8330-0011
Alice Sweeting 0000-0002-9185-6773
Emily Cust 0000-0001-6927-6329

Author contact details:

*** Corresponding author:**

Emily E. Cust
Email: emily.cust1@live.vu.edu.au
Institute for Health and Sport (IHES), Victoria University, P.O. Box 14428, Melbourne, VIC 8001, Australia. Western Bulldogs Football Club, Footscray, Melbourne, Australia

Alice J. Sweeting
Email: Alice.Sweeting@vu.edu.au
Institute for Health and Sport (IHES), Victoria University, P.O. Box 14428, Melbourne, VIC 8001, Australia. Western Bulldogs Football Club, Footscray, Melbourne, Australia

Kevin Ball
Email: Kevin.Ball@vu.edu.au
Institute for Health and Sport (IHES), Victoria University, P.O. Box 14428, Melbourne, VIC 8001, Australia.

Sam Robertson
Email: Sam.Robertson@vu.edu.au
Institute for Health and Sport (IHES), Victoria University, P.O. Box 14428, Melbourne, VIC 8001, Australia. Western Bulldogs Football Club, Footscray, Melbourne, Australia

Running title:

Machine and deep learning for sport movement recognition review

.

Abstract

Objective assessment of an athlete's performance is of importance in elite sports to facilitate detailed analysis. The implementation of automated detection and recognition of sport-specific movements overcomes the limitations associated with manual performance analysis methods. The object of this study was to systematically review the literature on machine and deep learning for sport-specific movement recognition using inertial measurement unit (IMU) and, or computer vision data inputs. A search of multiple databases was undertaken. Included studies must have investigated a sport-specific movement and analysed via machine or deep learning methods for model development. A total of 52 studies met the inclusion and exclusion criteria. Data pre-processing, processing, model development and evaluation methods varied across the studies. Model development for movement recognition were predominantly undertaken using supervised classification approaches. A kernel form of the Support Vector Machine algorithm was used in 53% of IMU and 50% of vision-based studies. Twelve studies used a deep learning method as a form of Convolutional Neural Network algorithm and one study also adopted a Long Short Term Memory architecture in their model. The adaptation of experimental set-up, data pre-processing, and model development methods are best considered in relation to the characteristics of the targeted sports movement(s).

Key Words:

Sport movement classification; inertial sensors; computer vision; machine learning; performance analysis.

1. Introduction

Performance analysis in sport science has experienced considerable recent changes, due largely to access to improved technology and increased applications from computer science. Manual notational analysis or coding in sports, even when performed by trained analysts, has limitations. Such methods are typically time intensive, subjective in nature, and prone to human error and bias. Automating sport movement recognition and its application towards coding has the potential to enhance both the efficiency and accuracy of sport performance analysis. The potential automation of recognising human movements, commonly referred to as human activity recognition (HAR), can be achieved through machine or deep learning model approaches. Common data inputs are obtained from inertial measurement units (IMUs) or vision. Detection refers to the identification of a targeted instance, i.e., tennis strokes within a continuous data input signal (Bulling, Blanke, & Schiele, 2014). Recognition or classification of movements involves further interpretations and labelled predictions of the identified instance (Bulling et al., 2014; Bux, Angelov, & Habib, 2017), i.e., differentiating tennis strokes as a forehand or backhand. In machine and deep learning, a model represents the statistical operations involved in the development of an automated prediction task (LeCun, Yoshua, & Geoffrey, 2015; Shalev-Shwartz & Ben-David, 2014).

Human activities detected by inertial sensing devices and computer vision are represented as wave signal features corresponding to specific actions, which can be logged and extracted. Human movement activities are considered hierarchically structured and can be broken down to basic movements. Therefore, the context of signal use, intra-class variability, and inter-class similarity between activities require consideration during experimental set-up and model development. Wearable IMUs contain a combination of accelerometer, gyroscope, and magnetometer sensors measuring along one to three axes. These sensors quantify acceleration, angular velocity, and the direction and orientation of travel respectively (Gastin, McLean, Breed, & Spittle, 2014). These sensors can capture repeated movement patterns during sport training and competitions (Camomilla, Bergamini, Fantozzi, & Vannozzi, 2018; Chambers, Gabbett, Cole, & Beard, 2015; J. F. Wagner, 2018). Advantages include being wireless, lightweight and self-contained in operation. Inertial measurement units have been utilised in quantifying physical output and tackling impacts in Australian Rules football (Gastin et al., 2014; Gastin, McLean, Spittle, & Breed, 2013) and rugby (Gabbett, Jenkins, & Abernethy, 2012, 2011; Howe, Aughey, Hopkins, Stewart, & Cavanagh, 2017; Hulin, Gabbett, Johnston, & Jenkins, 2017). Other applications include swimming analysis (Mooney, Corley, Godfrey, Quinlan, & ÓLaighin, 2015), golf swing kinematics (Lai, Hetchl, Wei, Ball, & McLaughlin, 2011), over-ground running speeds (Wixted, Billing, & James, 2010), full motions in alpine skiing (Yu et al., 2016); and the detection and evaluation of cricket bowling (McNamara, Gabbett, Blanch, & Kelly, 2017; McNamara, Gabbett, Chapman, Naughton, & Farhart, 2015; Wixted, Portus, Spratford, & James, 2011).

Computer vision has applications for performance analysis including player tracking, semantic analysis, and movement analysis (Stein et al., 2018; Thomas, Gade, Moeslund, Carr, & Hilton, 2017). Automated movement recognition approaches require several pre-processing steps including athlete detection and tracking, temporal cropping and targeted action recognition, which are dependent upon the sport and footage type (Barris & Button, 2008; Saba & Altameem, 2013; Thomas et al., 2017). Several challenges including occlusion, viewpoint variations, and environmental conditions may impact results, depending on the camera set-up (Poppe, 2010; Zhang et al., 2017). Developing models to automate sports-vision coding may improve resource efficiency and reduce feedback times. For example, coaches and athletes involved in time-intensive notational tasks, including post-swim race analysis, may benefit from rapid objective feedback before the next race in the event program (Liao, Liao, & Liu, 2003; Victor, He, Morgan, & Miniutti, 2017). For detecting and recognising movements, body worn sensor signals do not suffer from the same environmental constraints and stationary set-up of video cameras. Furthermore, multiple sensors located on different body segments have been argued to provide more specific signal representations of targeted movements (J. B. Yang, Nguyen, San, Li, & Shonali, 2015). But it is not clear if this is solely conclusive, and the use of body worn sensors in some sport competitions may be impractical or not possible.

Machine learning algorithms learn from data input for automated model building and perform tasks without being explicitly programmed. The algorithm goal is to output a response function $h\sigma(\vec{x})$ that will predict a ground truth variable y from an input vector of variables $\vec{x}$. Models are run for classification techniques to predict a target class (Kotsiantis, Zaharakis, & Pintelas, 2007), or regression to predict discrete or continuous values. Models are aimed at finding an optimal set of parameters σ to describe the response function, and then make predictions on unseen unlabelled data input. Within these, model training approaches can generally run as supervised learning, unsupervised learning or semi-supervised learning (Mohammed, Khan, & Bashier, 2016; Sze, Chen, Yang, & Emer, 2017).

Processing raw data is limited for conventional machine learning algorithms, as they are unable to effectively be trained on abstract and high-dimensional data that is inconsistent, contains missing values or noisy artefacts (Bux et al., 2017; Kautz, 2017). Consequently, several pre-processing stages are required to create a suitable data form for input into the classifier algorithm (Figo, Diniz, Ferreira, & Cardoso, 2010). Filtering (Figo et al., 2010; Wundersitz, Gastin, Robertson, Davey, & Netto, 2015), window capture durations (Mitchell, Monaghan, & O’Connor, 2013; Preece, Goulermas, Kenney, & Howard, 2009; Wundersitz, Josman, et al., 2015), and signal frequency cut-offs (Wundersitz, Gastin, Richter, Robertson, & Netto, 2015; Wundersitz, Gastin, Robertson, et al., 2015) are common techniques applied prior to data prior to dynamic human movement recognition. Well-established filters for processing motion signal data include the Kalman filter (Kautz, 2017; Titterton & Weston, 2009; D. Wagner, Kalischewski, Velten, & Kummert, 2017) and a Fourier transform filter (Preece, Goulermas, Kenney, Howard, et al., 2009) such as a fast Fourier transform (Kapela, Świetlicka, Rybarczyk, Kolanowski, & O’Connor, 2015; Preece, Goulermas, Kenney, & Howard, 2009). Near real-time processing benefits from reducing memory requirements, computational demands, and essential bandwidth during whole model implementation. Signal feature extraction and selection favours faster processing by reducing the signals to the critical features that can discriminate the targeted activities (Bulling et al., 2014). Feature extraction involves identifying the key features that help maximise classifier success, and removing features that have minimal impact in the model (Mannini & Sabatini, 2010). Thus, feature selection involves constructing data representations in subspaces with reduced dimensions. These identified variables are represented in a compact feature variable (Mannini & Sabatini, 2010). Common methods include principal component analysis (PCA) (Gløersen, Myklebust, Hallén, & Federolf, 2018; Young & Reinkensmeyer, 2014), vector coding techniques (Hafer & Boyer, 2017) and empirical cumulative distribution functions (ECDF) (Plötz, Hammerla, & Olivier, 2011). An ECDF approach has been shown to be advantageous over PCA as it derives representations of raw input independent of the absolute data ranges, whereas PCA is known to have reduced performance when the input data is not properly normalised (Plötz et al., 2011). For further detailed information on the acquisition, filtering and analysis of IMU data for sports application and vision-based human activity recognition, see (Kautz, 2017) and (Bux et al., 2017), respectively.

Deep learning is a division of machine learning, characterised by deeper neural network model architectures and are inspired by the biological neural networks of the human brain (Bengio, 2013; LeCun et al., 2015; Sze et al., 2017). The deeper hierarchical models create a profound architecture of multiple hidden layers based on representative learning with several processing and abstraction layers (Bux et al., 2017; J. B. Yang et al., 2015). These computational models allow data input features to be automatically extracted from raw data and transformed to handle unstructured data, including vision (LeCun et al., 2015; Ravi, Wong, Lo, & Yang, 2016). This direct input avoids several processing steps required in machine learning during training and testing, therefore reducing overall computational times. A current key element within deep learning is backpropagation (Hecht-Nielsen, 1989; LeCun, Bottou, Orr, & Müller, 1998). Backpropagation is a fast and computationally efficient algorithm, using gradient descent, that allows training deep neural networks to be tractable (Sze et al., 2017). Human activity recognition has mainly been performed using conventional machine learning classifiers. Recently, deep learning techniques have enhanced the bench mark and applications for IMUs (Kautz et al., 2017; Ravi et al., 2016; Ronao & Cho, 2016; J. B. Yang et al., 2015; Zebin, Scully, & Ozanyan, 2016; Zeng et al., 2014) and vision (Ji, Yang, Yu, & Xu, 2013; Karpathy et al., 2014a; Krizhevsky, Sutskever, & Hinton, 2012; Nibali, He, Morgan, & Greenwood, 2017) in human movement recognition producing more superior model performance accuracy.

The objective of this study was to systematically review the literature investigating sport-specific automated movement detection and recognition. The review focusses on the various technologies, analysis techniques and performance outcome measures utilised. There are several reviews within this field that are sensor-based including wearable IMUs for lower limb biomechanics and exercises (Fong & Chan, 2010; M. O’Reilly, Caulfield, Ward, Johnston, & Doherty, 2018), swimming analysis (Magalhaes, Vannozzi, Gatta, & Fantozzi, 2015; Mooney et al., 2015), quantifying sporting movements (Chambers et al., 2015) and physical activity monitoring (C. C. Yang & Hsu, 2010). A recent systematic review has provided an evaluation on the in-field use of inertial-based sensors for various performance evaluation applications (Camomilla et al., 2018). Vision-based methods for human activity recognition (Aggarwal & Xia, 2014; Bux et al., 2017; Ke et al., 2013; Zhang et al., 2017), semantic human activity recognition (Ziaeefard & Bergevin, 2015) and motion analysis in sport (Barris & Button, 2008) have also been reviewed. However, to date, there is no systematic review across sport-specific movement detection and recognition via machine or deep learning. Specifically, incorporating IMUs and vision-based data input, focussing on in-field applications as opposed to laboratory-based protocols and detailing the analysis and machine learning methods used.

Considering the growth in research and potential field applications, such a review is required to understand the research area. This review aims to characterise the evolving techniques and inform researchers of possible improvements in sports analysis applications. Specifically: 1) What is the current scope for IMUs and computer vision in sport movement detection and recognition? 2) Which methodologies, inclusive of signal processing and model learning techniques, have been used to achieve sport movement recognition? 3) Which evaluation methods have been used in assessing the performance of these developed models?

2. Methods

2.1 Search strategy

The preferred PRISMA recommendations (Moher, Liberati, Tetzlaff, Altman, & Group, 2009) for systematic reviews were used. A literature search was undertaken by the first author on the following databases; IEEE Xplore, PubMed, ScienceDirect, Scopus, Academic Search Premier, and Computer and Applied Science Complete. The searched terms were categorised in order to define the specific participants, methodology and evaluated outcome measure in-line with the review aims. Searches used a combination of key words with AND/OR phrases which are detailed in Table 1. Searches were filtered for studies from January 2000 to May 2018 as no relevant studies were identified prior to this. Further studies were manually identified from the bibliographies of database-searched studies identified from the abstract screen phase, known as snowballing. Table 2 provides the inclusion and exclusion criteria of this review.

*****Table 1 near here: Key word search term strings per database *****

*****Table 2 near here: Inclusion and exclusion criteria*****

2.2 Data extraction

The first author extracted and collated the relevant information from the full manuscripts identified for final review. A total of 18 parameters were extracted from the 52 research studies, including the title, author, year of publication, sport, participant details, sport movement target(s), device specifications, device sample frequency, pre-processing methods, processing methods, feature selected, feature extraction, machine learning model used, model evaluation, model performance accuracy, validation method, samples collected, and computational information. A customised Microsoft Excel^{TM} spreadsheet was developed to categorise the relevant extracted information from each study. Participant characteristics of number of participants, gender, and competition level, then if applicable a further descriptor specific to a sport, for example, 'medium-paced cricket bowler'. Athlete and participant experience level was categorised as written in the corresponding study to avoid misrepresentations. The age of participants was not considered an important characteristic required for model development. The individual ability in which the movement is performed accounts for the discriminative signal features associated with the movements. For the purposes of this review, a sport-specific movement was defined from a team or individual sport, and training activities associated with a particular sport. For example, weight-lifting as strength training, recognised under the Global Association of International Sports Federations. The targeted sports and specific movements were defined for either detection or recognition. Model development techniques used included pre-processing methods to transform data to a more suitable form for analysis, processing stages to segment data for identified target activities, feature extraction and selections techniques, and the learning algorithm(s). Model evaluation measures extracted were the model performance assessment techniques used, ground-truth validation comparison, number of data samples collected, and the model performance outcomes results reported. If studies ran multiple experiments using several algorithms, only the superior algorithm and relevant results were reported as the best method. This was done so in the interest of concise reporting to highlight favourable method approaches (Sprager & Juric, 2015). Any further relevant results or information identified from the studies was included as a special remark (Sprager & Juric, 2015). Hardware and specification information extracted included the IMU or video equipment used, number of units, attachment of sensors (IMUs), sample frequency, and sensor data types used in analysis (IMUs). Studies identified and full data extracted were reviewed by a second author.

3. Results

An outline of the search results and study exclusions has been provided in Fig 1. Of the initial database search which identified 4885 results, a final 52 studies met criteria for inclusion in this review. Of these, 29 used IMUs and 22 were vision-based. One study (Ó Conaire et al., 2010) used both sensors and vision for model development separately then together via data fusion. Tables 3 - 8 provide a description of the characteristics of the reviewed studies, detailed in the following sections.

***** Fig 1 near here: PRISMA flow diagram *****

3.1 Experimental design

A variety of sports and their associated sport-specific movements were investigated, implementing various experimental designs as presented in Tables 5 and 7. Across the studies, sports reported were tennis (n = 10), cricket (n = 3), weightlifting or strength training (n = 6), swimming (n = 4), skateboarding (n = 2), ski jumping (n = 2), snowboarding (n = 1), golf (n = 4), volleyball (n = 2), rugby (n = 2), ice hockey (n = 2), gymnastics (n = 2), karate (n = 1), basketball (n = 3), Gaelic football (n = 1), hurling (n = 1), boxing (n = 2), running (n = 2), diving (n = 1), squash (n = 1), badminton (n = 1), cross-country skiing (n = 2) and soccer (n = 4). The Sports 1-M dataset (Karpathy et al., 2014b) was also reported, which consists of 1,133,158 video URLs annotated automatically with 487 sport labels using the YouTube Topic API. A dominant approach was the classification of main characterising actions for each sport. For example, serve, forehand, backhand strokes in tennis (Connaghan et al., 2011; Kos & Kramberger, 2017; Ó Conaire et al., 2010; Shah, Chokalingam, Paluri, & Pradeep, 2007; Srivastava et al., 2015), and the four competition strokes in swimming (Jensen, Blank, Kugler, & Eskofier, 2016; Jensen, Prade, & Eskofier, 2013; Liao et al., 2003; Victor et al., 2017). Several studies further classified sub-categories of actions. For example, three further classes of the two main classified snowboarding trick types Grinds and Airs (Groh, Fleckenstein, & Eskofier, 2016), and further classifying the main tennis stroke types as either flat, topspin or slice (Srivastava et al., 2015). Semantic descriptors were reported for classification models that predicted athlete training background, experience and fatigue level. These included running (Buckley et al., 2017; Kobsar, Osis, Hettinga, & Ferber, 2014), rating of gymnastic routines (Reily, Zhang, & Hoff, 2017), soccer pass classification based on its quality (Horton, Gudmundsson, Chawla, & Estephan, 2014), cricket bowling legality (Qaisar et al., 2013; Salman, Qaisar, & Qamar, 2017), ski jump error analysis (Brock & Ohgi, 2017; Brock, Ohgi, & Lee, 2017) and strength training technique deviations (M. A. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017a; M. O'Reilly et al., 2015; M. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017). One approach (Yao & Fei-Fei, 2010), encoded the mutual context of human pose and sporting equipment using semantics, to facilitate the detection and classification of movements including a cricket bat and batsman coupled movements.

Total participant numbers for IMU-based studies ranged from one (Qaisar et al., 2013) to 30 (Kautz et al., 2017). Reported data individual instance sample sizes for sensor studies ranged from 150 (Salman et al., 2017) to 416, 737 (Rassem, El-Beltagy, & Saleh, 2017). Vision-based studies that explicitly reported total participant details ranged from five (Ó Conaire et al., 2010) to 40 (Victor et al., 2017). Vision dataset sample sizes varied across studies, from 50 individual action clips (Liao et al., 2003) to 15, 000 (Victor et al., 2017). One study (Karpathy et al., 2014a) used the publicly available Sports-1M, as previously described. Vision-based studies also reported datasets in total time, 10.3 hours (Bertasius, Park, Yu, & Shi, 2017), 3 hours (Montoliu, Martín-Félez, Torres-Sospedra, & Martínez-Usó, 2015), 1, 500 minutes (Shah et al., 2007), and 50 hours (Kapela et al., 2015), and by frame numbers, 6, 035 frames (Zhu, Xu, Gao, & Huang, 2006) and 10, 115 frames (Reily et al., 2017).

3.2 Inertial measurement unit specifications

A range of commercially available and custom-built IMUs were used in the IMU-based studies (n= 30), as presented in Table 3. Of these, 23% reported using a custom-built sensor. Of the IMU-based studies, the number of sensors mounted or attached to each participant or sporting equipment piece ranged from one to nine. The majority of studies (n= 22) provided adequate details of sensor specifications including sensor type, axes, measurement range, and sample rate used. At least one characteristic of sensor measurement range or sample rate used in data collection was missing from eight studies. All studies used triaxial sensors and collected accelerometer data. For analysis and model development, individual sensor data consisted of only accelerometer data (n = 8), both accelerometer and gyroscope data (n = 15), and accelerometer, gyroscope and magnetometer data (n = 7). The individual sensor measurement ranges reported for accelerometer were ± 1.5 g to ± 16 g, gyroscope ± 500 ◦/s to ± 2000 ◦/s, magnetometer ± 1200 µT or 1.2 to 4 Ga. Individual sensor sample rates ranged from 10 Hz to 1000 Hz for accelerometers, 10 Hz to 500 Hz for gyroscopes and 50 Hz to 500 Hz for magnetometers.

***** Table 3 near here*****

3.3 Vision capture specification

Several experimental set-ups and specifications were reported in the total 23 vision-based studies (Table 4). Modality was predominately red, green, blue (RGB) cameras. Depth cameras were utilised (Kasiri-Bidhendi, Fookes, Morgan, Martin, & Sridharan, 2015; Kasiri, Fookes, Sridharan, & Morgan, 2017; Reily et al., 2017), which add depth perception for 3-dimensional image mapping. Seven studies clearly reported the use of a single camera set-up (Couceiro, Dias, Mendes, & Araújo, 2013; Díaz-Pereira, Gómez-Conde, Escalona, & Olivieri, 2014; Hachaj, Ogiela, & Koptyra, 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Nibali et al., 2017; Reily et al., 2017). One study reported 16 stationary positioned cameras at a ‘bird’s eye view’ (Montoliu et al., 2015), and Ó Conaire et al. (2010) reported the use of one overhead and 8 stationary cameras around a tennis court baseline, although data from two cameras were only used in final analysis due to occlusion issues. Sample frequency and, or pixel resolution were reported in seven of the studies (Couceiro et al., 2013; Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Montoliu et al., 2015; Victor et al., 2017; Zhu et al., 2006), with sample frequencies ranging from 30 Hz to 210 Hz.

***** Table 4 near here*****

3.4 Inertial measurement unit recognition model development methods

Key stages of model development from data pre-processing to recognition techniques for IMU-based studies are presented in Table 5. Data pre-processing filters were reported as either a low-pass filter (n = 7) (Adelsberger & Tröster, 2013; Buckley et al., 2017; Kelly, Coughlan, Green, & Caulfield, 2012; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2015, 2017; Rindal, Seeberg, Tjønnås, Haugnes, & Sandbakk, 2018), high-pass filter (n = 2) (Kautz et al., 2017; Schuldhaus et al., 2015), or calibration with a filter (Salman et al., 2017). Processing methods were reported in 67% of the IMU-based studies (Adelsberger & Tröster, 2013; Anand, Sharma, Srivastava, Kaligounder, & Prakash, 2017; Brock et al., 2017; Buckley et al., 2017; Buthe, Blanke, Capkevics, & Tröster, 2016; Groh et al., 2016; Groh, Fleckenstein, Kautz, & Eskofier, 2017; Groh, Kautz, & Schuldhaus, 2015; Jensen et al., 2016, 2015; Jiao, Wu, Bie, Umek, & Kos, 2018; Kautz et al., 2017; Kobsar et al., 2014; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2017; Ó Conaire et al., 2010; Pernek, Kurillo, Stiglic, & Bajcsy, 2015; Qaisar et al., 2013; Salman et al., 2017; Schuldhaus et al., 2015). Methods included, calibration of data (Groh et al., 2016, 2017; Jensen et al., 2015; Qaisar et al., 2013), a one-second window centred around identified activity peaks in the signal (Adelsberger & Tröster, 2013; Schuldhaus et al., 2015), temporal alignment (Pernek et al., 2015), normalisation (Ó Conaire et al., 2010), outlier adjustment (Kobsar et al., 2014) or removal (Salman et al., 2017), and sliding windows ranging from one to 3.5 seconds across the data (Jensen et al., 2016). The three studies that investigated trick classification in skateboarding (Groh et al., 2017, 2015) and snowboarding (Groh et al., 2016) corrected data for different rider board stance styles, termed Regular or Goofy, by inverting signal axes.

Movement detection methods were specifically reported in 16 studies (Adelsberger & Tröster, 2013; Anand et al., 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015, Jensen et al., 2013, 2015; Kautz et al., 2017; Kelly et al., 2012; Kos & Kramberger, 2017; Ó Conaire et al., 2010; Rindal et al., 2018; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside, Cant, Connolly, & Reid, 2017). Detection methods included thresholding (n = 5), windowing segmenting (n = 4), and a combination of threshold and windowing techniques (n = 5).

Signal feature extraction techniques were reported in 80% of the studies, with the number of feature parameters in a vector ranging from a vector of normalised X, Y, Z accelerometer signals (Ó Conaire et al., 2010) to 240 features (M. A. O'Reilly et al., 2017a). Further feature selection to reduce the dimensionality of the feature vector was used in 11 studies. Both feature extraction and selection methods varied considerably across the literature (Table 5).

Algorithms trialled for movement recognition were diverse across the literature (Table 5). Supervised classification using a kernel form of Support Vector Machine (SVM) was most prevalent (n = 16) (Adelsberger & Tröster, 2013; Brock & Ohgi, 2017; Brock et al., 2017; Buckley et al., 2017; Buthe et al., 2016; Groh et al., 2016, 2017, 2015; Jensen et al., 2016; Kautz et al., 2017; Kelly et al., 2012; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside et al., 2017). The next highest tested were Naïve Bayesian (NB) (n = 8) (Buckley et al., 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015; Kautz et al., 2017; Salman et al., 2017; Schuldhaus et al., 2015) and k-Nearest Neighbour (kNN) (n = 8) (Buckley et al., 2017; Groh et al., 2016, 2017, 2015; Kautz et al., 2017; Ó Conaire et al., 2010; Salman et al., 2017; Whiteside et al., 2017), followed by Random Forests (RF) (n = 7) (Buckley et al., 2017; Groh et al., 2017; Kautz et al., 2017; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2017; Salman et al., 2017; Whiteside et al., 2017). Supervised learning algorithms were the most common (n = 29). One study used an unsupervised discriminative analysis approach for detection and classification of tennis strokes (Kos & Kramberger, 2017). Five IMU-based study investigated a deep learning approach including using Convolutional Neural Networks (CNN) (Anand et al., 2017; Brock et al., 2017; Jiao et al., 2018; Kautz et al., 2017; Rassem et al., 2017) and Long Short Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) architectures (Rassem et al., 2017; Sharma, Srivastava, Anand, Prakash, & Kaligounder, 2017). In order to assess the effectiveness of the various classifiers from each study, model performance measures quantify and visualise the predictive performance as reported in the following section.

***** Table 5 near here*****

3.5 Inertial measurement unit recognition model evaluation

Reported performance evaluations of developed models across the IMU-based studies are shown in Table 6. Classification accuracy, as a percentage score for the number of correct predictions by total number of predictions made, was the main model evaluation measure (n = 24). Classification accuracies across studies ranged between 52% (Brock & Ohgi, 2017) to 100% (Buckley et al., 2017). Generally, the reported highest accuracy for a specific movement was ≥ 90% (n = 17) (Adelsberger & Tröster, 2013; Anand et al., 2017; Buckley et al., 2017; Connaghan et al., 2011; Groh et al., 2015; Jensen et al., 2013; Jiao et al., 2018; Kobsar et al., 2014; Kos & Kramberger, 2017; M. A. O'Reilly et al., 2017a; Ó Conaire et al., 2010; Pernek et al., 2015; Qaisar et al., 2013; Rindal et al., 2018; Schuldhaus et al., 2015; Srivastava et al., 2015; Whiteside et al., 2017) and ≥ 80% to 90% (n = 7) (Brock & Ohgi, 2017; Brock et al., 2017; Groh et al., 2017; Jensen et al., 2016; M. O'Reilly et al., 2015, 2017; Salman et al., 2017). As an estimate of the generalised performance of a trained model on $n - x$ samples, a form of leave-one-out cross validation (LOO-CV) was used in 47% of studies (Buthe et al., 2016; Groh et al., 2016, 2017, 2015, Jensen et al., 2016, 2013; Kobsar et al., 2014; M. O'Reilly et al., 2015, 2017; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015). Precision, specificity and sensitivity (also referred to as recall) evaluations were derived for detection (n = 6) and classification models (n = 10). Visualisation of prediction results in the form of a confusion matrix featured in six studies (Buthe et al., 2016; Groh et al., 2017; Kautz et al., 2017; Pernek et al., 2015; Rindal et al., 2018; Whiteside et al., 2017).

***** Table 6 near here*****

3.6 Vision recognition model development methods

Numerous processing and recognition methods featured across the vision-based studies to transform and isolated relevant input data (Table 7). Pre-processing stages were reported in 14 of studies, and another varied 13 studies also provided details of processing techniques. Signal feature extraction and feature selection methods used were reported in 78% of studies.

Both machine (n = 16) and deep learning (n = 7) algorithms were used to recognise movements from vision data. Of these, a kernel form of the SVM algorithm was most common in the studies (n = 10) (Couceiro et al., 2013; Horton et al., 2014; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Li et al., 2018; Montoliu et al., 2015; M. A. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017b; Ó Conaire et al., 2010; Reily et al., 2017; Shah et al., 2007; Zhu et al., 2006). Other algorithms included kNN (n = 3) (Díaz-Pereira et al., 2014; Montoliu et al., 2015; Ó Conaire et al., 2010), decision tree (DT) (n = 2) (Kapela et al., 2015; Liao et al., 2003), RF (n = 2) (Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017), and Multilayer Perceptron (MLP) (n = 2) (Kapela et al., 2015; Montoliu et al., 2015). Deep learning was investigated in seven studies (Bertasius et al., 2017; Ibrahim, Muralidharan, Deng, Vahdat, & Mori, 2016; Karpathy et al., 2014a; Nibali et al., 2017; Ramanathan et al., 2015; Tora, Chen, & Little, 2017; Victor et al., 2017) of which used CNNs or LSTM RNNs as the core model structure.

***** Table 7 near here*****

3.7 Vision recognition model evaluation

Performance evaluation methods and results for vision-based studies are reported in Table 8. As with IMU-based studies, classification accuracy was the common method for model evaluations, featured in 61%. Classification accuracies were reported between 60.9% (Karpathy et al., 2014a) and 100% (Hachaj et al., 2015; Nibali et al., 2017). In grouping the reported highest accuracies for a specific movement that were ≥ 90% (n = 9) (Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Li et al., 2018; Montoliu et al., 2015; Nibali et al., 2017; Ó Conaire et al., 2010; Reily et al., 2017; Shah et al., 2007), and ≥ 80% to 90% (n = 2) (Horton et al., 2014; Yao & Fei-Fei, 2010). A confusion matrix as a visualisation of model prediction results was used in nine studies (Couceiro et al., 2013; Hachaj et al., 2015; Ibrahim et al., 2016; Karpathy et al., 2014a; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Lu, Okuma, & Little, 2009; Shah et al., 2007; Tora et al., 2017). Two studies assessed and reported their model computational average speed (Lu et al., 2009) and time (Reily et al., 2017).

***** Table 8 near here*****

4 Discussion

The aim of this systematic review was to evaluate the use of machine and deep learning for sport-specific movement recognition from IMUs and, or computer vision data inputs. Overall, the search yielded 52 studies, categorised as 29 which used IMUs, 22 vision-based and one study using both IMUs and vision. Automation or semi-automated sport movement recognition models working in near-real time is of particular interest to avoid the error, cost and time associated with manual methods. Evident in the literature, models are trending towards the potential to provide optimised objective assessments of athletic movement for technical and tactical evaluations. The majority of studies achieved favourable movement recognition results for the main characterising actions of a sport, with several studies exploring further applications such as an automated skill quality evaluation or judgement scoring, for example automated ski jump error evaluation (Brock et al., 2017).

Experimental set-up of IMU placement and numbers assigned per participant varied between sporting actions. The sensor attachment locations set by researchers appeared dependent upon the specific sporting conditions and movements, presumably to gain optimal signal data. Proper fixation and alignment of the sensor axes with limb anatomical axes is important in reducing signal error (Fong & Chan, 2010). The attachment site hence requires a biomechanical basis for accuracy of the movement being targeted to obtain reliable data. Single or multiple sensor use per person also impacts model development trade-off between accuracy, analysis complexity, and computational speed or demands. In tennis studies, specificity whilst using a single sensor was demonstrated by mounting the IMU on the wrist or forearm of the racquet arm (Connaghan et al., 2011; Kos & Kramberger, 2017; Srivastava et al., 2015; Whiteside et al., 2017). A single sensor may also be mounted in a low-profile manner on sporting equipment (Groh et al., 2016, 2017, 2015; Jensen et al., 2015). Unobtrusive use of a single IMU to capture generalised movements across the whole body was demonstrated, with an IMU mounted on the posterior head in swimming (Jensen et al., 2016, 2013), lower back during running (Kobsar et al., 2014), and between the shoulder blades in rugby union (Kelly et al., 2012).

The majority of vision-based studies opted for a single camera set-up of RGB modality. Data output from a single camera as opposed to multiple minimises the volume of data to process, therefore reducing computational effort. However, detailed features may go uncaptured, particularly in team sport competition which consists of multiple individuals participating in the capture space at one time. In contrast, a multiple camera set-up reduces limitations including occlusion and viewpoint variations. However, this may also increase the complexity of the processing and model computational stages. Therefore, a trade-off between computational demands and movement recording accuracy often needs to be made. As stated earlier, the placement of cameras needs to suit the biomechanical nature of the targeted movement and the environment situated in. Common camera capture systems used in sports science research such as Vicon Nexus (Oxford, UK) and OptiTrack (Oregon, USA) were not present in this review. As this review targeted studies investigating during on-field or in-situation sporting contexts, efficiency in data collection is key for routine applications in training and competition. A simple portable RGB camera is easy to set-up in a dynamic and changing environment, such as different soccer pitches, rather than a multiple capture system such as Vicon that requires calibrated precision and are substantially more expensive.

Data acquisition and type from an IMU during analysis appears to influence model trade-off between accuracy and computational effort of performance. The use of accelerometer, gyroscope or magnetometer data may depend upon the movement properties analysed. Within tennis studies, gyroscope signals were the most efficient at discriminating between stroke types (Buthe et al., 2016; Kos & Kramberger, 2017) and detecting an athlete's fast feet court actions (Buthe et al., 2016). In contrast, accelerometer signals produced higher classification accuracies in classifying tennis stroke skills levels (Connaghan et al., 2011). The authors expected lower gyroscope classification accuracies as temporal orientation measures between skill levels of tennis strokes will differ (Connaghan et al., 2011). Conversely, data fusion from all three individual sensors resulted in a more superior model for classifying advanced, intermediate and novices tennis player strokes (Connaghan et al., 2011). Fusion of accelerometer and vision data also resulted in a higher classification accuracy for tennis stroke recognition (Ó Conaire et al., 2010).

Supervised learning approaches were dominant across IMU and vision-based studies. This is a method which involves a labelled ground truth training dataset typically manually annotated by sport analysts. Labelled data instances were recorded as up to 15, 000 for vision-based (Victor et al., 2017) and 416, 737 for sensor-based (Rassem et al., 2017) studies. Generation of a training data set for supervised learning can be a tedious and labour-intensive task. It is further complicated if multiple sensors or cameras are incorporated for several targeted movements. A semi-supervised or unsupervised learning approach may be advantageous as data labelling is minimal or not required, potentially reducing human errors in annotation. An unsupervised approach could suit specific problems to explain key data features, via clustering (Mohammed et al., 2016; Sze et al., 2017). Results computed by an unsupervised model (Kos, Ženko, Vlaj, & Kramberger, 2016) for tennis serve, forehand and backhand stroke classification compared favourbaly well against a proposed supervised approach (Connaghan et al., 2011).

Recognition of sport-specific movements was primarily achieved using conventional machine learning approaches, however nine studies implemented deep learning algorithms. It is expected that future model developments will progressively feature deep learning approaches due to development of better hardware, and the advantages of more efficient model learning on large data inputs (Sze et al., 2017). Convolutional Neural networks (CNN) (LeCun, Bottou, Bengio, & Haffner, 1998) were the core structure of five of the seven deep learning study models. Briefly, convolution applies several filters, known as kernels, to automatically extract features from raw data inputs. This process works under four key ideas to achieve optimised results: local connection, shared weights, pooling and applying several layers (LeCun et al., 2015; J. B. Yang et al., 2015). Machine learning classifiers modelled with generic hand-crafted features, were compared against a CNN for classifying nine beach volleyball actions using IMUs (Kautz et al., 2017). Unsatisfactory results were obtained from the machine learning model, and the CNN markedly achieved higher classification accuracies (Kautz et al., 2017). The CNN model produced the shortest overall computation times, requiring less computational effort on the same hardware (Kautz et al., 2017). Vision-based CNN models have also shown favourable results when compared to a machine learning study baseline (Karpathy et al., 2014a; Nibali et al., 2017; Victor et al., 2017). Specifically, consistency between a swim stroke detection model for continuous videos in swimming which was then applied to tennis strokes with no domain-specific settings introduced (Victor et al., 2017). The authors of this training approach (Victor et al., 2017) anticipate that this could be applied to train separate models for other sports movement detection as the CNN model demonstrated the ability to learn to process continuous videos into a 1-D signal with the signal peaks corresponding to arbitrary events. General human activity recognition using CNN have shown to be a superior approach over conventional machine learning algorithms using both IMUs

(Ravi et al., 2016; J. B. Yang et al., 2015; Zebin et al., 2016; Zeng et al., 2014; Zheng, Liu, Chen, Ge, & Zhao, 2014) and computer vision (Ji et al., 2013; Krizhevsky et al., 2012; LeCun et al., 2015). As machine learning algorithms extract heuristic features requiring domain knowledge, this creates shallower features which can make it harder to infer high-level and context aware activities (J. B. Yang et al., 2015). Given the previously described advantages of deep learning algorithms which apply to CNN, and the recent results of deep learning, future model developments may benefit from exploring these methods in comparison to current bench mark models.

Model performance outcome metrics quantify and visualise the error rate between the predicted outcome and true measure. Comparatively, a kernel form of an SVM was the most common classifier implemented and produced the strongest machine learning approach model prediction accuracies across both IMU (Adelsberger & Tröster, 2013; Brock & Ohgi, 2017; Buthe et al., 2016; Groh et al., 2016, 2017, 2015; Jensen et al., 2016; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside et al., 2017) and vision-based study designs (Horton et al., 2014; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Li et al., 2018; Reily et al., 2017; Shah et al., 2007; Zhu et al., 2006). Classification accuracy was the most common reported measure followed by confusion matrices, as ways to clearly present prediction results and derive further measures of performance. Further measures included sensitivity (also called recall), specificity and precision, whereby results closer to 1.0 indicate superior model performance, compared to 0.0 or poor model performance. The F1-score (also called a F-measure or F-score) conveys the balances between the precision and sensitivity of a model. An in-depth analysis performance metrics specific to human activity recognition is located elsewhere (Minnen, Westeyn, Starner, Ward, & Lukowicz, 2006; Ward, Lukowicz, & Gellersen, 2011). Use of specific evaluation methods depends upon the data type. Conventional performance measures of error rate are generally unsuitable for models developed from skewed training data (Provost & Fawcett, 2001). Using conventional performance measures in this context will only take the default decision threshold on a model trained, if there is an uneven class distribution this may lead to imprecision (Provost & Fawcett, 2001; Seiffert, Khoshgoftaar, Van Hulse, & Napolitano, 2008). Alternative evaluators including Receiver Operating Characteristics (ROC) curves and its single numeric measure, Area Under ROC Curve (AUC), report model performances across all decision thresholds (Seiffert et al., 2008). Making evaluations between study methodology have inherent complications due to each formulating their own experimental parameter settings, feature vectors and training algorithms for movement recognition. The No-Free-Lunch theorems are important deductions in the formation of models for supervised machine learning (David H. Wolpert, 1996), and search and optimisation algorithms (D H Wolpert & Macready, 1997). The theorems broadly reference that there is no 'one model' that will perform optimally across all recognition problems. Therefore, experiments with multiple model development methods for a particular problem is recommended. The use of prior knowledge about the task should be implemented to adapt the model input and model parameters in order to improve overall model success (Shalev-Shwartz & Ben-David, 2014).

Acquisition of athlete specific information, including statistics on number, type and intensity of actions, may be of use in the monitoring of athlete load. Other potential applications include personalised movement technique analysis (M. O'Reilly et al., 2017), automated performance evaluation scoring (Reily et al., 2017) and team ball sports pass quality rating (Horton et al., 2014). However, one challenge lies in delivering consistent, individualised models across team field sports that are dynamic in nature. For example, classification of soccer shots and passes showed a decline in model performance accuracy from a closed environment to a dynamic match setting (Schuldhaus et al., 2015). A method to overcome accuracy limitations in dynamic team field sports associated with solely using IMUs or vision may be to implement data fusion (Ó Conaire et al., 2010). Furthermore, vision and deep learning approaches have demonstrated the ability to track and classify team sport collective court activities and individual player specific movements in volleyball (Ibrahim et al., 2016), basketball (Ramanathan et al., 2015) and ice hockey (Tora et al., 2017). Accounting for methods from experimental set-up to model evaluation, previous reported models should be considered and adapted based on the current problem. Furthermore, the balance between model computational efficiency, results accuracy and complexity trade-offs calculations are an important factor.

In the present study, meta-analysis was considered however variability across developed model parameter reporting and evaluation methods did not allow for this to be undertaken. As this field expands and further methodological approaches are investigated, it would be practical to review analysis approaches both within and between sports. This review was delimited to machine and deep learning approaches to sport movement detection and recognition. However, statistical or parametric approaches not considered here such as discriminative functional analysis may also show efficacy for sport-specific movement recognition. However, as the field of machine learning is a rapidly developing area shown to produce superior results, a review encompassing all possible other methods may have complicated the reporting. Since sport-specific movements and their environments alter the data acquisition and analysis, the sports and movements reported in the present study provide an overview of the current field implementations.

5 Conclusions

This systematic review reported on the literature using machine and deep learning methods to automate sport-specific movement recognition. In addressing the research questions, both IMUs and computer vision have demonstrated capacity in improving the information gained from sport movement and skill recognition for performance analysis. A range of methods for model development were used across the reviewed studies producing varying results. Conventional machine learning algorithms such as Support Vector Machines and Neural Networks were most commonly implemented. Yet in those studies which applied deep learning algorithms such as Convolutional Neural Networks, these methods outperformed the machine learning algorithms in comparison. Typically, the models were evaluated using a leave-one-out cross validation method and reported model performances as a classification accuracy score. Intuitively, the adaptation of experimental set-up, data processing, and recognition methods used are best considered in relation to the characteristics of the sport and targeted movement(s). Consulting current models within or similar to the targeted sport and movement is of benefit to address bench mark model performances and identify areas for improvement. The application within the sporting domain of machine learning and automated sport analysis coding for consistent uniform usage appears currently a challenging prospect, considering the dynamic nature, equipment restrictions and varying environments arising in different sports.

Future work may look to adopt, adapt and expand on current models associated with a specific sports movement to work towards flexible models for mainstream analysis implementation. Investigation of deep learning methods in comparison to conventional machine learning algorithms would be of particular interest to evaluate if the trend of superior performances is beneficial for sport-specific movement recognition. Analysis as to whether IMUs and vision alone or together yield enhanced results in relation to a specific sport and its implementation efficiency would also be of value. In consideration of the reported study information, this review can assist future researchers in broadening investigative approaches for sports performance analysis as a potential to enhancing upon current methods.

Acknowledgements

Not applicable.

Disclosure statement

No potential conflict of interest was reported by the authors. However, the last author is co-author of three studies included in this systematic review.

Funding

The authors received no specific funding for this work.

Author ORCID

Sam Robertson 0000-0002-8330-0011
Alice Sweeting 0000-0002-9185-6773
Emily Cust 0000-0001-6927-6329

References

Adelsberger, R., & Tröster, G. (2013). Experts lift differently: Classification of weight-lifting athletes. In *2013 IEEE International Conference on Body Sensor Networks* (pp. 1–6). Cambridge, MA: Body Sensor Networks (BSN). https://doi.org/10.1109/BSN.2013.6575458

Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3D data: A review. *Pattern Recognition Letters*, *48*, 70–80. https://doi.org/10.1016/j.patrec.2014.04.011

Anand, A., Sharma, M., Srivastava, R., Kaligounder, L., & Prakash, D. (2017). Wearable motion sensor based analysis of swing sports. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 261–267). https://doi.org/10.1109/ICMLA.2017.0-149

Barris, S., & Button, C. (2008). A review of vision-based motion analysis in sport. *Sports Medicine*, *38*(12), 1025–1043. https://doi.org/10.2165/00007256-200838120-00006

Bengio, Y. (2013). Deep learning of representations: Looking forward. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7978 LNAI*, 1–37. https://doi.org/10.1007/978-3-642-39593-2_1

Bertasius, G., Park, H. S., Yu, S. X., & Shi, J. (2017). Am I a baller? Basketball performance assessment from first-person videos. *Proceedings of the IEEE International Conference on Computer Vision*, 2196–2204. https://doi.org/10.1109/ICCV.2017.239

Brock, H., & Ohgi, Y. (2017). Assessing motion style errors in ski jumping using inertial sensor devices. *IEEE Sensors Journal*, (99), 1–11. https://doi.org/10.1109/JSEN.2017.2699162

Brock, H., Ohgi, Y., & Lee, J. (2017). Learning to judge like a human: convolutional networks for classification of ski jumping errors. *Proceedings of the 2017 ACM International Symposium* *on Wearable Computers - ISWC '17*, 106–113. https://doi.org/10.1145/3123021.3123038

Buckley, C., O'Reilly, M. A., Whelan, D., Vallely Farrell, A., Clark, L., Longo, V., … Caulfield, B. (2017). Binary classification of running fatigue using a single inertial measurement unit. In *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks* (pp. 197–201). IEEE. https://doi.org/10.1109/BSN.2017.7936040

Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, *46*(3), 1–33. https://doi.org/http://dx.doi.org/10.1145/2499621

Buthe, L., Blanke, U., Capkevics, H., & Tröster, G. (2016). A wearable sensing system for timing analysis in tennis. In *BSN 2016 - 13th Annual Body Sensor Networks Conference* (pp. 43–48). San Francisco, CA. https://doi.org/10.1109/BSN.2016.7516230

Bux, A., Angelov, P., & Habib, Z. (2017). Vision based human activity recognition: A review. In P. Angelov, A. Gegov, C. Jayne, & Q. Shen (Eds.), *Advances in Computational Intelligence Systems: Contributions Presented at the 16th UK Workshop on Computational Intelligence* (pp. 341–371). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-46562-3_23

Camomilla, V., Bergamini, E., Fantozzi, S., & Vannozzi, G. (2018). Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: a systematic review. *Sensors*, *18*(3), 873. https://doi.org/10.3390/s18030873

Chambers, R., Gabbett, T., Cole, M. H., & Beard, A. (2015). The use of wearable microsensors to quantify sport-specific movements. *Sports Medicine*, *45*(7), 1065–1081. https://doi.org/10.1007/s40279-015-0332-9

Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., Walsh, M., & O'Mathuna, C. (2011). Multi-sensor classification of tennis strokes. *Journal of IEEE Sensors*, 1437–1440.

Couceiro, M. S., Dias, G., Mendes, R., & Araújo, D. (2013). Accuracy of pattern detection methods in the performance of golf putting. *Journal of Motor Behavior*, *45*(1), 37–53. https://doi.org/10.1080/00222895.2012.740100

Díaz-Pereira, M. P., Gómez-Conde, I., Escalona, M., & Olivieri, D. N. (2014). Automatic recognition and scoring of olympic rhythmic gymnastic movements. *Human Movement Science*, *34*(1), 63–80. https://doi.org/10.1016/j.humov.2014.01.001

Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. M. P. (2010). Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, *14*(7), 645–662. https://doi.org/10.1007/s00779-010-0293-9

Fong, D. T.-P., & Chan, Y.-Y. (2010). The use of wearable inertial motion sensors in human lower limb biomechanics studies: A systematic review. *Sensors*, *10*(12), 11556–11565. https://doi.org/10.3390/s101211556

Gabbett, T., Jenkins, D., & Abernethy, B. (2012). Physical demands of professional rugby league training and competition using microtechnology. *Journal of Science and Medicine in Sport*, *15*, 80–86. https://doi.org/10.1016/j.jsams.2011.07.004

Gabbett, T., Jenkins, D. G., & Abernethy, B. (2011). Physical collisions and injury in professional rugby league match-play. *Journal of Science and Medicine in Sport*, *14*, 210–215. https://doi.org/10.1016/j.jsams.2011.01.002

Gastin, P. B., McLean, O. C., Breed, R. V., & Spittle, M. (2014). Tackle and impact detection in elite Australian football using wearable microsensor technology. *Journal of Sports Sciences*, *32*(10), 947–953. https://doi.org/10.1080/02640414.2013.868920

Gastin, P. B., McLean, O. C., Spittle, M., & Breed, R. V. (2013). Quantification of tackling demands in professional Australian football using integrated wearable athlete tracking technology. *Journal of Science and Medicine in Sport*, *16*(6), 589–593. https://doi.org/10.1016/j.jsams.2013.01.007

Gløersen, Ø., Myklebust, H., Hallén, J., & Federolf, P. (2018). Technique analysis in elite athletes using principal component analysis. *Journal of Sports Sciences*, *36*(2), 229–237. https://doi.org/10.1080/02640414.2017.1298826

Groh, B. H., Fleckenstein, M., & Eskofier, B. M. (2016). Wearable trick classification in freestyle snowboarding. In *13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)* (pp. 89–93). IEEE. https://doi.org/10.1109/BSN.2016.7516238

Groh, B. H., Fleckenstein, M., Kautz, T., & Eskofier, B. M. (2017). Classification and visualization of skateboard tricks using wearable sensors. *Pervasive and Mobile Computing*, *40*, 42–55. https://doi.org/10.1016/j.pmcj.2017.05.007

Groh, B. H., Kautz, T., & Schuldhaus, D. (2015). IMU-based trick classification in skateboarding.

In *KDD Workshop on Large-Scale Sports Analytics*.

Hachaj, T., Ogiela, M. R., & Koptyra, K. (2015). Application of assistive computer vision methods to Oyama karate techniques recognition. *Symmetry*, *7*(4), 1670–1698. https://doi.org/10.3390/sym7041670

Hafer, J. F., & Boyer, K. A. (2017). Variability of segment coordination using a vector coding technique: reliability analysis for treadmill walking and running. *Gait and Posture, 51*, 222–227. https://doi.org/10.1016/j.gaitpost.2016.11.004

Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. *Proceedings Of The International Joint Conference On Neural Networks*, *1*, 593–605. https://doi.org/10.1109/IJCNN.1989.118638

Hochreiter, S., & Schmidhuber, J. J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1–32. https://doi.org/10.1162/neco.1997.9.8.1735

Horton, M., Gudmundsson, J., Chawla, S., & Estephan, J. (2014). Classification of passes in football matches using spatiotemporal data. *ArXiv Preprint ArXiv:1407.5093*. https://doi.org/10.1145/3105576

Howe, S. T., Aughey, R. J., Hopkins, W. G., Stewart, A. M., & Cavanagh, B. P. (2017). Quantifying important differences in athlete movement during collision-based team sports: Accelerometers outperform global positioning systems. In *2017 IEEE International Symposium on Inertial Sensors and Systems* (pp. 1–4). Kauai, HI, USA: IEEE. https://doi.org/10.1109/ISISS.2017.7935655

Hulin, B. T., Gabbett, T., Johnston, R. D., & Jenkins, D. G. (2017). Wearable microtechnology can accurately identify collision events during professional rugby league match-play. *Journal of Science and Medicine in Sport*, *20*(7), 638–642. https://doi.org/http://dx.doi.org/10.1016/j.jsams.2016.11.006

Ibrahim, M., Muralidharan, S., Deng, Z., Vahdat, A., & Mori, G. (2016). A Hierarchical Deep Temporal Model for Group Activity Recognition. *Cvpr*, 1971–1980. https://doi.org/10.1109/CVPR.2016.217

Jensen, U., Blank, P., Kugler, P., & Eskofier, B. M. (2016). Unobtrusive and energy-efficient swimming exercise tracking using on-node processing. *IEEE Sensors Journal*, *16*(10), 3972–3980. https://doi.org/10.1109/JSEN.2016.2530019

Jensen, U., Prade, F., & Eskofier, B. M. (2013). Classification of kinematic swimming data with emphasis on resource consumption. In *2013 IEEE International Conference on Body Sensor Networks, BSN 2013*. https://doi.org/10.1109/BSN.2013.6575501

Jensen, U., Schmidt, M., Hennig, M., Dassler, F. A., Jaitner, T., & Eskofier, B. M. (2015). An IMU-based mobile system for golf putt analysis. *Sports Engineering*, *18*(2), 123–133. https://doi.org/10.1007/s12283-015-0171-9

Ji, S., Yang, M., Yu, K., & Xu, W. (2013). 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 221–231. https://doi.org/10.1109/TPAMI.2012.59

Jiao, L., Wu, H., Bie, R., Umek, A., & Kos, A. (2018). Multi-sensor Golf Swing Classification Using Deep CNN. *Procedia Computer Science*, *129*, 59–65. https://doi.org/10.1016/j.procs.2018.03.046

Kapela, R., Świetlicka, A., Rybarczyk, A., Kolanowski, K., & O'Connor, N. E. (2015). Real-time event classification in field sport videos. *Signal Processing: Image Communication, 35*, 35–45. https://doi.org/10.1016/j.image.2015.04.005

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014a). Large-scale video classification with convolutional neural networks. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference On*, 1725–1732. https://doi.org/10.1109/CVPR.2014.223

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014b). Large-scale video classification with convolutional nural networks. Retrieved December 18, 2017, from http://cs.stanford.edu/people/karpathy/deepvideo/

Kasiri-Bidhendi, S., Fookes, C., Morgan, S., Martin, D. T., & Sridharan, S. (2015). Combat sports analytics: Boxing punch classification using overhead depth imagery. In *2015 IEEE International Conference on Image Processing (ICIP)* (pp. 4545–4549). Quebec City, Canada: IEEE. https://doi.org/10.1109/ICIP.2015.7351667

Kasiri, S., Fookes, C., Sridharan, S., & Morgan, S. (2017). Fine-grained action recognition of boxing punches from depth imagery. *Computer Vision and Image Understanding*, *159*, 143–153. https://doi.org/10.1016/j.cviu.2017.04.007

Kautz, T. (2017). Acquisition, filtering and analysis of positional and inertial data in sports. *FAU Studies in Computer Science*, *2*.

Kautz, T., Groh, B. H., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B. M. (2017). Activity recognition in beach volleyball using a deep convolutional neural network. *Data Mining and Knowledge Discovery*, 1–28. https://doi.org/10.1007/s10618-017-0495-0

Ke, S. R., Thuc, H., Lee, Y. J., Hwang, J. N., Yoo, J. H., & Choi, K. H. (2013). A review on video-based human activity recognition. *Computers*, *2*, 88–131. https://doi.org/10.3390/computers2020088

Kelly, D., Coughlan, G. F., Green, B. S., & Caulfield, B. (2012). Automatic detection of collisions in elite level rugby union using a wearable sensing device. *Sports Engineering*, *15*(2), 81–92. Retrieved from https://0-link-springer-com.library.vu.edu.au/article/10.1007%2Fs12283-012-0088-5

Kobsar, D., Osis, S. T., Hettinga, B. A., & Ferber, R. (2014). Classification accuracy of a single tri-axial accelerometer for training background and experience level in runners. *Journal of Biomechanics*, *47*(10), 2508–2511. https://doi.org/10.1016/j.jbiomech.2014.04.017

Kos, M., & Kramberger, I. (2017). A wearable device and system for movement and biometric data Acquisition for sports applications. *IEEE Access*, 1–1. https://doi.org/10.1109/ACCESS.2017.2675538

Kos, M., Ženko, J., Vlaj, D., & Kramberger, I. (2016). Tennis stroke detection and classification using miniature wearable IMU device. In *International Conference on Systems, Signals, and Image Processing*. https://doi.org/10.1109/IWSSIP.2016.7502764

Kotsiantis, S., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, *31*, 501–520. https://doi.org/10.1115/1.1559160

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances In Neural Information Processing Systems*, 1097–1105. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

Lai, D. T. H., Hetchl, M., Wei, X., Ball, K., & McLaughlin, P. (2011). On the difference in swing arm kinematics between low handicap golfers and non-golfers using wireless inertial sensors. *Procedia Engineering*, *13*, 219–225. https://doi.org/10.1016/j.proeng.2011.05.076

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

LeCun, Y., Bottou, L., Orr, G. B., & Müller, K. R. (1998). Efficient backprop. In *Neural Networks: Tricks of the Trade* (Vol. 1524, pp. 9–50).

LeCun, Y., Yoshua, B., & Geoffrey, H. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, J., Tian, Q., Zhang, G., Zheng, F., Lv, C., & Wang, J. (2018). Research on hybrid information recognition algorithm and quality of golf swing. *Computers and Electrical Engineering*, 1–13. https://doi.org/10.1016/j.compeleceng.2018.02.013

Liao, W. H., Liao, Z. X., & Liu, M. J. (2003). Swimming style classification from video sequences. In Kinmen (Ed.), *16th IPPR Conference on Computer Vision, Graphics and Image Processing* (pp. 226–233). ROC.

Lu, W. L., Okuma, K., & Little, J. J. (2009). Tracking and recognizing actions of multiple hockey players using the boosted particle filter. *Image and Vision Computing*, *27*(1–2), 189–205. https://doi.org/10.1016/j.imavis.2008.02.008

Magalhaes, F. A. de, Vannozzi, G., Gatta, G., & Fantozzi, S. (2015). Wearable inertial sensors in swimming motion analysis: A systematic review. *Journal of Sports Sciences*, *33*(7), 732–745. https://doi.org/10.1080/02640414.2014.962574

Mannini, A., & Sabatini, A. M. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, *10*(2), 1154–1175. https://doi.org/10.3390/s100201154

McNamara, D. J., Gabbett, T., Blanch, P., & Kelly, L. (2017). The relationship between wearable microtechnology device variables and cricket fast bowling intensity. *International Journal of Sports Physiology and Performance*, 1–20. https://doi.org/https://doi.org/10.1123/ijspp.2016-0540

McNamara, D. J., Gabbett, T., Chapman, P., Naughton, G., & Farhart, P. (2015). The validity of microsensors to automatically detect bowling events and counts in cricket fast bowlers. *International Journal of Sports Physiology and Performance*, *10*(1), 71–75. https://doi.org/10.1123/ijspp.2014-0062

Minnen, D., Westeyn, T. L., Starner, T., Ward, J. a, & Lukowicz, P. (2006). Performance metrics and evaluation issues for continuous activity recognition. In *Proc. Int. Workshop on Performance Metrics for Intelligent Systems* (pp. 141–148). https://doi.org/10.1145/1889681.1889687

Mitchell, E., Monaghan, D., & O’Connor, N. E. (2013). Classification of sporting activities using smartphone accelerometers. *Sensors (Basel, Switzerland)*, *13*(4), 5317–5337. https://doi.org/10.3390/s130405317

Mohammed, M., Khan, M., & Bashier, E. (2016). *Machine Learning: Algorithms and Applications*. Milton: CRC Press.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Montoliu, R., Martín-Félez, R., Torres-Sospedra, O., & Martínez-Usó, A. (2015). Team activity recognition in Association football using a bag-of-words-based method. *Human Movement Science*, *41*, 165–178. https://doi.org/10.1016/j.humov.2015.03.007

Mooney, R., Corley, G., Godfrey, A., Quinlan, L. R., & ÓLaighin, G. (2015). Inertial sensor technology for elite swimming performance analysis: A systematic review. *Sensors*, *16*(1), 18. https://doi.org/10.3390/s16010018

Nibali, A., He, Z., Morgan, S., & Greenwood, D. (2017). Extraction and classification of diving clips from continuous video footage. *ArXiv*, *pre-print*. Retrieved from https://arxiv.org/pdf/1705.09003.pdf

O’Reilly, M. A., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017a). Classification of lunge biomechanics with multiple and individual inertial measurement units. *Sports Biomechanics*, *16*(3), 342–360. https://doi.org/10.1080/14763141.2017.1314544

O’Reilly, M. A., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017b). Technology in strength and conditioning tracking lower-limb exercises with wearable sensors. *Journal of Strength and Conditioning Research*, *31*(6), 1726–1736.

O’Reilly, M., Caulfield, B., Ward, T., Johnston, W., & Doherty, C. (2018). Wearable Inertial Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review. *Sports Medicine*. https://doi.org/10.1007/s40279-018-0878-4

O’Reilly, M., Whelan, D., Chanialidis, C., Friel, N., Delahunt, E., Ward, T., & Caulfield, B. (2015). Evaluating squat performance with a single inertial measurement unit. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks*. IEEE. https://doi.org/10.1109/BSN.2015.7299380

O’Reilly, M., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017). Classification of deadlift biomechanics with wearable inertial measurement units. *Journal of Biomechanics*, *58*, 155–161. https://doi.org/10.1080/14763141.2017.1314544

Ó Conaire, C., Connaghan, D., Kelly, P., O’Connor, N. E., Gaffney, M., & Buckley, J. (2010). Combining inertial and visual sensing for human action recognition in tennis. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams* (pp. 51–56). ACM. https://doi.org/10.1145/1877868.1877882

Pernek, I., Kurillo, G., Stiglic, G., & Bajcsy, R. (2015). Recognizing the intensity of strength training exercises with wearable sensors. *Journal of Biomedical Informatics*, *58*, 145–155. https://doi.org/10.1016/j.jbi.2015.09.020

Plötz, T., Hammerla, N. Y., & Olivier, P. (2011). Feature learning for activity recognition in ubiquitous computing. *International Joint Conference on Artificial Intelligence (IJCAI)*, 1729.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing*, *28*(6), 976–990. https://doi.org/10.1016/j.imavis.2009.11.014

Preece, S. J., Goulermas, J. Y., Kenney, L., & Howard, D. (2009). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering*, *56*(3), 871–879. https://doi.org/10.1109/TBME.2008.2006190

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors: A review of classification techniques. *Physiological Measurement*, *30*(4), R1–R33. https://doi.org/10.1088/0967-3334/30/4/R01

Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, *42*(3), 203–231. https://doi.org/10.1023/A:1007601015854

Qaisar, S., Imtiaz, S., Glazier, P., Farooq, F., Jamal, A., Iqbal, W., & Lee, S. (2013). A method for cricket bowling action classification and analysis using a system of inertial sensors. In

International Conference on Computational Science and its Applications (pp. 396–412). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39649-6

Ramanathan, V., Huang, J., Abu-El-Haija, S., Gorban, A., Murphy, K., & Fei-Fei, L. (2015). Detecting events and key actors in multi-person videos. https://doi.org/10.1109/CVPR.2016.332

Rassem, A., El-Beltagy, M., & Saleh, M. (2017). Cross-country skiing gears classification using deep learning. *ArXiv Preprint ArXiv:1706.08924*. Retrieved from https://arxiv.org/pdf/1706.08924v1.pdf

Ravi, D., Wong, C., Lo, B., & Yang, G.-Z. (2016). A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE Journal of Biomedical and Health Informatics*, *21*(1), 1–1. https://doi.org/10.1109/JBHI.2016.2633287

Reily, B., Zhang, H., & Hoff, W. (2017). Real-time gymnast detection and performance analysis with a portable 3D camera. *Computer Vision and Image Understanding*, *159*, 154–163. https://doi.org/10.1016/j.cviu.2016.11.006

Rindal, O. M. H., Seeberg, T. M., Tjønnås, J., Haugnes, P., & Sandbakk, Ø. (2018). Automatic classification of sub-techniques in classical cross-country skiing using a machine learning algorithm on micro-sensor data. *Sensors (Switzerland)*, *18*(1), 75. https://doi.org/10.3390/s18010075

Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications*, *59*, 235–244. https://doi.org/10.1016/j.eswa.2016.04.032

Saba, T., & Altameem, A. (2013). Analysis of vision based systems to detect real time goal events in soccer videos. *Applied Artificial Intelligence*, *27*(7), 656–667. https://doi.org/10.1080/08839514.2013.787779

Salman, M., Qaisar, S., & Qamar, A. M. (2017). Classification and legality analysis of bowling action in the game of cricket. *Data Mining and Knowledge Discovery*, *31*(6), 1706–1734. https://doi.org/10.1007/s10618-017-0511-4

Schuldhaus, D., Zwick, C., Körger, H., Dorschky, E., Kirk, R., & Eskofier, B. M. (2015). Inertial sensor-based approach for shot/ pass classification during a soccer match. In *Proc. 21st ACM KDD Workshop on Large-Scale Sports Analytics* (pp. 1–4). Sydney, Australia.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2008). RUSBoost: Improving classification performance when training data is skewed. In *9th International Conference on Pattern Recognition* (pp. 1–4). https://doi.org/10.1109/ICPR.2008.4761297

Shah, H., Chokalingam, P., Paluri, B., & Pradeep, N. (2007). Automated stroke classification in tennis. *Image Analysis and Recognition*, 1128–1137.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: from theory to algorithms*. New York, USA: Cambridge University Press.

Sharma, M., Srivastava, R., Anand, A., Prakash, D., & Kaligounder, L. (2017). Wearable motion sensor based phasic analysis of tennis serve for performance feedback. In *2017 IEEE International Conference on Acoustics, Speech and Siginal Processing* (pp. 5945–5949). New Orleans, LA: IEEE.

Sprager, S., & Juric, M. B. (2015). *Inertial sensor-based gait recognition: A review*. *Sensors (Switzerland)* (Vol. 15). https://doi.org/10.3390/s150922089

Srivastava, R., Patwari, A., Kumar, S., Mishra, G., Kaligounder, L., & Sinha, P. (2015). Efficient characterization of tennis shots and game analysis using wearable sensors data. In *2015 IEEE Sensors- Proceedings* (pp. 1–4). Busan. https://doi.org/10.1109/ICSENS.2015.7370311

Stein, M., Janetzko, H., Lamprecht, A., Breitkreutz, T., Zimmermann, P., Goldlücke, B., … Keim, D. A. (2018). Bring it to the pitch: combining video and movement data to enhance team sport analysis. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 13–22. https://doi.org/10.1109/TVCG.2017.2745181

Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. (2017). Efficient processing of deep neural networks: A tutorial and survey. *IEEE*, *105*(2), 2295–2329. Retrieved from http://arxiv.org/abs/1703.09039

Thomas, G., Gade, R., Moeslund, T. B., Carr, P., & Hilton, A. (2017). Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, *159*, 3–18. https://doi.org/10.1016/j.cviu.2017.04.011

Titterton, D. H., & Weston, J. L. (2009). *Strapdown inertial navigation technology* (2nd ed.). Reston, VA: AIAA.

Tora, M. R., Chen, J., & Little, J. J. (2017). Classification of puck possession events in ice hockey.

In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 147–154). https://doi.org/10.1109/CVPRW.2017.24

Victor, B., He, Z., Morgan, S., & Miniutti, D. (2017). Continuous video to simple signals for swimming stroke detection with convolutional neural networks. *ArXiv Preprint ArXiv:1705.09894*. https://doi.org/10.1111/j.1467-8330.1974.tb00606.x

Wagner, D., Kalischewski, K., Velten, J., & Kummert, A. (2017). Activity recognition using inertial sensors and a 2-D convolutional neural network. In IEEE (Ed.), *2017 10th International Workshop on Multidimensional (nD) Systems (nDS)* (pp. 1–6). https://doi.org/10.1109/NDS.2017.8070615

Wagner, J. F. (2018). About motion measurement in sports based on gyroscopes and accelerometers - an engineering point of view. *Gyroscopy and Navigation*, *9*(1), 1–18. https://doi.org/10.1134/S2075108718010091

Ward, J. A., Lukowicz, P., & Gellersen, H.-W. (2011). Performance metrics for activity recognition. In *ACM Trans. on Intelligent Systems and Technology* (Vol. 2, pp. 111–132).

Whiteside, D., Cant, O., Connolly, M., & Reid, M. (2017). Monitoring hitting load in tennis using inertial sensors and machine learning. *International Journal of Sports Physiology and Performance*, 1–20. https://doi.org/https://doi.org/10.1123/ijspp.2016-0683

Wixted, A., Billing, D. C., & James, D. A. (2010). Validation of trunk mounted inertial sensors for analysing running biomechanics under field conditions, using synchronously collected foot contact data. *Sports Engineering*, *12*(4), 207–212. https://doi.org/10.1007/s12283-010-0043-2

Wixted, A., Portus, M., Spratford, W., & James, D. A. (2011). Detection of throwing in cricket using wearable sensors. *Sports Technology*, *4*(3–4), 134–140. https://doi.org/10.1080/19346182.2012.725409

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, *8*(7), 1341–1390. https://doi.org/10.1162/neco.1996.8.7.1391

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimisation. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82. https://doi.org/10.1023/A:1021251113462

Wundersitz, D. W., Gastin, P. B., Richter, C., Robertson, S., & Netto, K. J. (2015). Validity of a trunk-mounted accelerometer to assess peak accelerations during walking, jogging and running. *European Journal of Sport Science*, *15*(5), 382–390. https://doi.org/10.1080/17461391.2014.955131

Wundersitz, D. W., Gastin, P. B., Robertson, S., Davey, P. C., & Netto, K. J. (2015). Validation of a trunk-mounted accelerometer to measure peak impacts during team sport movements. *International Journal of Sports Medicine*, *36*(9), 742–746. https://doi.org/10.1055/s-0035-1547265

Wundersitz, D. W., Josman, C., Gupta, R., Netto, K. J., Gastin, P. B., & Robertson, S. (2015). Classification of team sport activities using a single wearable tracking device. *Journal of Biomechanics*, *48*(15), 3975–3981. https://doi.org/10.1016/j.jbiomech.2015.09.015

Yang, C. C., & Hsu, Y. L. (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, *10*(8), 7772–7788. https://doi.org/10.3390/s100807772

Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., & Shonali, K. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of the 24th International Conference on Artificial Intelligence* (pp. 3995–4001).

Yao, B., & Fei-Fei, L. (2010). Modeling mutual context of object and human pose in human-object interaction activities. In *Computer Vision and Pattern Recognition* (pp. 17–24). IEEE.

Young, C., & Reinkensmeyer, D. J. (2014). Judging complex movement performances for excellence: a principal components analysis-based technique applied to competitive diving. *Human Movement Science*, *36*, 107–122. https://doi.org/10.1016/j.humov.2014.05.009

Yu, G., Jang, Y. J., Kim, J., Kim, J. H., Kim, H. Y., Kim, K., & Panday, S. B. (2016). Potential of IMU sensors in performance analysis of professional alpine skiers. *Sensors (Switzerland)*, *16*(4), 1–21. https://doi.org/10.3390/s16040463

Zebin, T., Scully, P. J., & Ozanyan, K. B. (2016). Human Activity Recognition with Inertial Sensors Using a Deep Learning Approach. *Proc. of IEEE Sensors 2016*, (1), 1–3. https://doi.org/10.1109/ICSENS.2016.7808590

Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014). Convolutional neural networks for human activity recognition using mobile sensors. In *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services* (pp. 197–205). https://doi.org/10.4108/icst.mobicase.2014.257786

Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., & Li, Z. (2017). A review on human activity recognition using vision-based method. *Journal of Healthcare Engineering*, *2017*, 1–31. https://doi.org/10.1155/2017/3090343

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). Time series classification using multi-channels deep convolutional neural networks. In *International Conference on Web-Age Information Management* (pp. 298–310). Springer. https://doi.org/10.1007/978-3-319-08010-9_33

Zhu, G., Xu, C., Gao, W., & Huang, Q. (2006). Action recognition in broadcast tennis video. *Computer Vision in Human-Computer Interaction*, 89–98. https://doi.org/10.1007/11754336_9

Ziaeefard, M., & Bergevin, R. (2015). Semantic human activity recognition: A literature review. *Pattern Recognition*, *48*(8), 2329–2345. https://doi.org/10.1016/j.patcog.2015.03.006

1040

Figure 1 - PRISMA flow diagram

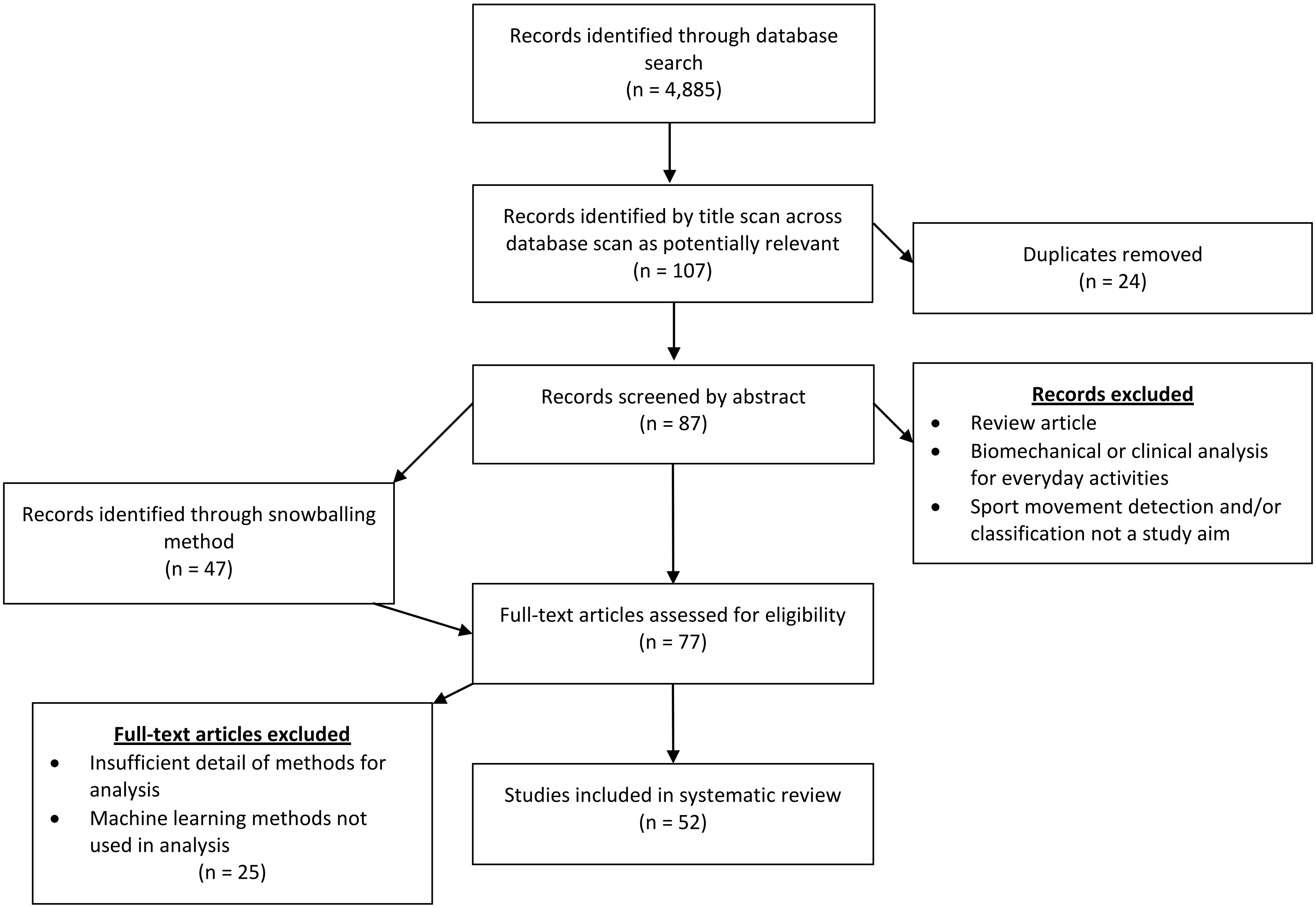

Figure 1 PRISMA flow diagram for study search, screen and selection process.

Table 1 Database key word searches.

Table 1. Key word search term strings per database.

Database key word searches
IEEE Xplore: ((((inertial sensor OR accelerometer OR gyroscope OR IMU OR microsensor)) AND (sport OR athlete* OR match OR game OR training)) AND (detection OR recognition OR classification)) AND (movement OR skill) ((((sport OR athlete* OR player*)) AND (video OR vision)) AND movement classification)
PubMed: ((((inertial sensor OR accelerometer OR gyroscope OR IMU OR microsensor)) AND (sport OR athlete* OR match OR game OR training)) AND (detection OR recognition OR classification)) AND (movement OR skill) ((((((((Vision OR video OR camera OR footage OR computer vision)) AND (sport OR athlete* OR match OR game OR training)) AND (detection OR recognition OR classification)) AND (movement OR skill))) AND human) NOT clinical)) NOT review
ScienceDirect: ((sport OR athlete* OR player*)) and ((inertial sensor OR accelerometer) ((sport OR athlete* OR player*)) and TITLE-ABSTR-KEY((vision OR video OR camera) AND (detection OR classification)).
Scopus: ((((inertial sensor OR accelerometer OR gyroscope OR IMU OR microsensor)) AND (sport OR athlete* OR match OR game OR training)) AND (detection OR recognition OR classification)) AND (movement OR skill) ((((sport OR athlete* OR player*)) AND (video OR vision)) AND movement classification)
Academic Search Premier: ((((inertial sensor OR accelerometer OR gyroscope OR IMU OR microsensor)) AND (sport OR athlete* OR match OR game OR training)) AND (detection OR recognition OR classification)) AND (movement OR skill) ((((sport OR athlete* OR player*)) AND (video OR vision)) AND movement classification)
Computer and Applied Science Complete: ((((inertial sensor OR accelerometer OR gyroscope OR IMU OR microsensor)) AND (sport OR athlete* OR match OR game OR training)) AND (detection OR recognition OR classification)) AND (movement OR skill) ((((Vision OR video OR camera OR footage OR computer vision)) AND (sport OR athlete* OR match OR game OR training)) AND (detection OR recognition OR classification)) AND (movement OR skill)

* Entails truncation, i.e., finding all terms that begin with the string of text written before it.

Table 2 Study inclusion and exclusion criteria

Table 2 Study inclusion and exclusion criteria.

Inclusion criteria	**Exclusion criteria**
• Original peer reviewed published manuscripts • Aimed at a sport-specific movement or skill, • Used IMUs and/or computer vision input datasets for model development • Investigated as an in-field application of the technology to the sporting movement • Defined clear data processing and model development methods inclusive of machine or deep learning algorithms for semi-automated or automated movement recognition • Published as full-length studies written in English	• Solely investigated gait analysis for clinical purposes • Solely investigated every day or non-sport-specific locomotion i.e., walking downstairs • Solely investigated player field positional tracking methods using data such as X, Y coordinates or displacement without any form of sport-specific skill detection and classification associated to it • Used ball trajectory and audio cue data as the major determinant for event detection • Data collection conducted within a laboratory setting under controlled protocol • Data processing pipelines or recognition model development methodology not clearly defined • Review studies

Table 3 Inertial measurement unit specifications.

Reference	**Sensor model**	**Sensor No.**	**Sensor placement**	**Accelerometer**			**Gyroscope**			**Magnetometer**		
				Axes	**Range**	**Sample rate**	**Axes**	**Range**	**Sample rate**	**Axes**	**Range** (1 Ga = 100 µT)	**Sample rate**
(Adelsberger & Tröster, 2013)	Ethos	3	Left ankle, wrist, lower back	3	± 6 g	NR	3	± 2000 ◦/s	NR	3	4 Ga	NR
(Anand, Sharma, Srivastava, Kaligounder, & Prakash, 2017)	Samsun Gear 2 smart watch	1	Wrist of hitting hand	3	± 8 g	100 Hz	3	± 2000 ◦/s	100 Hz			
(Brock & Ohgi, 2017)	Logical Product SS-WS1215/SS-WS1216, Fukuoka, Japan	9	Pelvis, right and left thighs, right and left shanks, right and left upper arms, both ski blades above the boot	3	± 5 g (body) ± 16 g (ski)	500 Hz	3	± 1500 ◦/s	500 Hz	3	± 1.2 Gauss full-scale	500 Hz
(Brock, Ohgi, & Lee, 2017)	Logical Product SS-WS1215/SS-WS1216, Fukuoka, Japan	9	Pelvis, right and left thighs, right and left shanks, right and left ski anterior to ski binding, right and left upper arm	3	± 5 g (body) ± 16 g (ski)	500 Hz	3	± 1500 ◦/s	500 Hz	3	± 1.2 Gauss full-scale	500 Hz
(Buckley et al., 2017)	Shimmer3 (Realtime Technologies Ltb. Dublin, Ireland)	3	Right and left shanks 2cm above lateral malleolus, 5th lumbar spinous process	3	± 8 g	256 Hz	3	± 1000 ◦/s	256 Hz	3	± 4 Gauss full-scale	256 Hz
(Buthe, Blanke, Capkevics, & Tröster, 2016)	EXLs33 IMU	3	Tennis racquet, on each shoe	3	± 16 g	200 Hz	3	± 500 ◦/s	200 Hz	3	NR	200 Hz
(Connaghan et al., 2011)	Custom Tyndall developed TennisSense WIMU system	1	Forearm of racquet arm	3	NR	NR	3	NR	NR	3	NR	NR

Table 3 continued.

Reference	Sensor model	Sensor No.	Sensor placement	Accelerometer			Gyroscope			Magnetometer		
				Axes	**Range**	**Sample rate**	**Axes**	**Range**	**Sample rate**	**Axes**	**Range** (1 Ga = 100 µT)	**Sample rate**
(Groh, Kautz, & Schuldhaus, 2015)	miPod sensor system	1	Underside of skateboard on the right side of front axis.	3	± 16g	200 Hz	3	± 2000 ◦/s	200 Hz	3	± 1200 µT	200 Hz
(Groh, Fleckenstein, & Eskofier, 2016)	miPod sensor system	1	Top of snowboard behind the front binding	3	± 16 g	200 Hz	3	± 2000 ◦/s	200 Hz	3	± 1200 µT	200 Hz
(Groh, Fleckenstein, Kautz, & Eskofier, 2017)	miPod sensor system	1	Underside of skateboard on the right side of front axis.	3	± 16 g	200 Hz	3	± 2000 ◦/s	200 Hz	3	± 1200 µT	200 Hz
(Jiao, Wu, Bie, Umek, & Kos, 2018)	NR	2	Golf club (location not specified)	3	NR	NR	3	NR	NR			
(Jensen et al., 2015)	Shimmer™ 2R sensor nodes (Realtime	1	Golf club head	3	± 1.5 g	256 Hz	3	± 500 ◦/s	256 Hz	NR	NR	NR
(Jensen, Blank, Kugler, & Eskofier, 2016)	Shimmer™ 2R sensor nodes (Realtime Technologies Ltb. Dublin, Ireland)	1	Back of head under a swim cap	3	± 1.5 g	10.24 Hz to 204.8 Hz	3	± 500 ◦/s	10.24 Hz to 204.8 Hz	NR	NR	NR
(Jensen, Prade, & Eskofier, 2013)	Shimmer™ (Realtime Technologies Ltb. Dublin, Ireland)	1	Back of head above swim cap	3	± 1.5 g	200 Hz	3	± 500 ◦/s	200 Hz	NR	NR	NR
(Kautz et al., 2017)	Bosch BMA280	1	Wrist of dominant hand	3	± 16 g	39 Hz	NR	NR	NR	NR	NR	NR
(Kelly, Coughlan, Green, & Caulfield, 2012)	SPI Pro	1	Between the shoulder blades	3	NR	39 Hz	NR	NR	NR	NR	NR	NR

Table 3 continued.

Reference	**Sensor model**	**Sensor No.**	**Sensor placement**	**Accelerometer**			**Gyroscope**			**Magnetometer**		
				Axes	**Range**	**Sample rate**	**Axes**	**Range**	**Sample rate**	**Axes**	**Range** (1 Ga = 100 µT)	**Sample rate**
(Kobsar, Osis, Hettinga, & Ferber, 2014)	G-Link wireless accelerometer node (Microstrain Inc., VT)	1	Lower back on the L3 vertebra region	3	± 10 g	617 Hz	NR	NR	NR	NR	NR	NR
(Kos & Kramberger, 2017)	Custom sensor	1	Wrist of racquet arm	3	± 16 g	NR	3	± 2000 ◦/s	NR	NR	NR	NR
(Ó Conaire et al., 2010)	Custom sensor	6	Left and right wrists, left and right ankles, chest, lower back	3	± 12 g	120 Hz	NR	NR	NR	NR	NR	NR
(O'Reilly et al., 2015)	Shimmer™ sensor (Realtime Technologies Ltb. Dublin, Ireland)	1	5th lumbar vertebra	3	± 16 g	51.2 Hz	3	± 500 ◦/s	51.2 Hz	3	± 1 Ga	51.2 Hz
(O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017a)	Shimmer™ sensor (Realtime Technologies Ltb. Dublin, Ireland)	5	5th lumbar vertebra, mid-point on right and left thighs, right and left shanks 2cm above lateral malleolus	3	± 2 g	51.2 Hz	3	± 500 ◦/s	51.2 Hz	3	± 1.9 Ga	51.2 Hz
(O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017b)	Shimmer™ sensor (Realtime Technologies Ltb. Dublin, Ireland)	5	Spinous process of the fifth lumbar vertebra, mid-point of both femurs, right and left shanks 2 cm above the lateral malleolus	3	± 2 g	51.2 Hz	3	± 500 ◦/s	51.2 Hz	3	± 1.9 Ga	51.2 Hz
(Pernek, Kurillo, Stiglic, & Bajcsy, 2015)	Custom sensor	5	Chest, left and right wrists, left and right upper arms	3	NR	30 Hz	NR	NR	NR	NR	NR	NR
(Qaisar et al., 2013)	Custom sensor	3	Bowling arm: upper arm, elbow joint, wrist	3	NR	150 Hz	3	NR	150 Hz	NR	NR	NR

Table 3 continued.

Reference	**Sensor model**	**Sensor No.**	**Sensor placement**	**Accelerometer**			**Gyroscope**			**Magnetometer**		
				Axes	**Range**	**Sample rate**	**Axes**	**Range**	**Sample rate**	**Axes**	**Range** (1 Ga = 100 µT)	**Sample rate**
(Rassem, El-Beltagy, & Saleh, 2017)	NR	1	NR	3	NR	50 Hz						
(Rindal, Seeberg, Tjønnås, Haugnes, & Sandbakk, 2018)	IsenseU Move+	2	Chest, Lower arm	3	NR	20 Hz	3	NR	20 Hz			
(Salman, Qaisar, & Qamar, 2017)	Custom sensor	3	Bowling arm: upper arm, forearm, wrist	3	NR	150 Hz	3	NR	150 Hz	NR	NR	NR
(Schuldhaus et al., 2015)	Custom sensor	2	Cavity of each shoe	3	± 16g	1000 Hz	NR	NR	NR	NR	NR	NR
(Srivastava et al., 2015)	Samsung Gear S smart watch	1	Wrist of racquet arm	3	± 8 g	25 Hz	3	± 2000 ◦/s	25 Hz	NR	NR	NR
(Whiteside, Cant, Connolly, & Reid, 2017)	IMeasureU IMU (Auckland, New Zealand)	1	Wrist of racquet arm	3	± 16 g	500 Hz	3	± 2000 ◦/s	500 Hz	3	± 1200 µT	500 Hz

g G-forces, *Ga* gauss, *Hz* Hertz, *IMU* inertial measurement unit, *µT* micro Tesla
NR not reported: study either did not directly report the specification or the device did not include the sensor type

Table 4 Vision-based camera specifications.

Reference	Camera model	Modality	Camera No.	Data collection setting
(Bertasius, Park, Yu, & Shi, 2017)	GoPro Hero 3 Black Edition	RGB	1	100 fps 1280 x 960 pixels
(Couceiro, Dias, Mendes, & Araújo, 2013)	Casio Exilim - High Speed EX-FH25. Focal length lens of 26 mm	RGB	1	Resolution 480 x 360 pixels 210 Hz
(Díaz-Pereira, Gómez-Conde, Escalona, & Olivieri, 2014)	Sony Handycam DCR-SR78	RGB	1	
(Hachaj, Ogiela, & Koptyra, 2015)	Kinetic 2 SDK system	3 Dimensional	1	30 Hz
(Horton, Gudmundsson, Chawla, & Estephan, 2014)	NR	NR	NR	NR
(Ibrahim, Muralidharan, Deng, Vahdat, & Mori, 2016)	NR	NR	NR	NR
(Kapela, Świetlicka, Rybarczyk, Kolanowski, & O'Connor, 2015)	NR	NR	NR	NR
(Karpathy et al., 2014)	NR	NR	NR	NR
(Kasiri-Bidhendi, Fookes, Morgan, Martin, & Sridharan, 2015)	Swisse-range SR4000 time-of-flight (MESA Imaging AG, Switzerland)	Depth Camera at 5 m overhead height	1	25 fps 176 x 144 pixels
(Kasiri, Fookes, Sridharan, & Morgan, 2017)	Swisse-range SR4000 time-of-flight (MESA Imaging AG, Switzerland)	Depth Camera at 5 m overhead height	1	25 fps 176 x 144 pixels
(Li et al., 2018)	iPhone5s, 6, 6plus, 6s, 7	RGB	1	30 fps
(Liao, Liao, & Liu, 2003)	NR	RGB	NR	NR
(Lu, Okuma, & Little, 2009)	NR	RGB	NR	NR

Table 4 continued.

Reference	**Camera model**	**Modality**	**Camera No.**	**Data collection setting**
(Montoliu, Martín-Félez, Torres-Sospedra, & Martínez-Usó, 2015)	NR	NR	16 synchronized and stationary with a 'bird's eye view' positioned along a soccer pitch	25 fps
(Nibali, He, Morgan, & Greenwood, 2017)	NR	RGB	One fixed	NR
(Ó Conaire et al., 2010)	IP camera	RGB	One overhead and eight around court baseline positioned	NR
(Ramanathan et al., 2015)	NR	NR	NR	NR
(Reily, Zhang, & Hoff, 2017)	Kinetic 2	Depth Camera	1	NR
(Shah, Chokalingam, Paluri, & Pradeep, 2007)	NR	RGB	NR	NR
(Tora, Chen, & Little, 2017)	NR	NR	NR	NR
(Victor, He, Morgan, & Miniutti, 2017)	NR	RGB	NR	Swimming: 50 fps Tennis: 30 fps
(Yao & Fei-Fei, 2010)	NR	RGB	NR	NR
(Zhu, Xu, Gao, & Huang, 2006)	Live Broadcast vision	RGB	NR	Video compressed in MPEG-2 standard with a frame resolution 352 x 288 pixels

fps frames per second, *Hz* hertz, *MPEG* Moving Picture Experts Group, *RGB* red green blue
NR not reported: study either did not directly report the specification or the device did not include the sensor type

Table 5 Inertial measurement unit study description and model characteristics.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset sample No.**	**Data pre-processing**			**Feature extraction**	**Feature selection**	**Recognition algorithm**
				Filter	**Processing**	**Detection**			
(Adelsberger & Tröster, 2013)	Weight-lifting: thruster (squat press)	16: four females and 12 males, beginner to expert		Low-pass filter	1 s window	Heuristically found threshold value to derive start and end indices of each thruster episode	Accelerometer magnitude modelled on sum of six Gaussian functions with four parameters each: scale αi, amplitude offset βi, standard deviation σi, and mean value μi	1.5 s window around detected signal peaks. Nelder Mead simplex direct search MATLAB	SVM
(Anand, Sharma, Srivastava, Kaligounder, & Prakash, 2017)	Tennis: forehand topspin, forehand slice, backhand topspin, backhand slice, serve Badminton: serve, clear, drop, smash Squash: forehand, backhand, serve	31 tennis players, 34 badminton players, 5 squash players	Total training set: ~8500. Total testing set: ~ 7100			Detection shot: 3 cues to identify shot regions across the three sports: 1) threshold, 2) jerk based detection, 3) shot shape-based detection. Once shot swing detected a fixed number or sample before and after impact point assigned as shot region	Seven shot windows developed for each stage of a shot. Three feature set types generated from all shot windows resulting in ~2000 features including: 1) statistical features, 2) pairwise correlation coefficients between elements of the window set, 3) shape-based features	Pearson correlation coefficient minimum redundancy maximum relevance (MRMR) technique	LR, bi-directional LSTM
(Brock & Ohgi, 2017)	Ski Jumping: error jump, non-error jump	Four: male, junior athletes					Set 1: discrete feature values based on one-dimensional data points built from the raw and processed data of every sensor Set 2: different time-series features based on the estimated positions and orientations of every sensor		SVM, DTW

Table 5 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset sample No.**	**Data pre-processing**			**Feature extraction**	**Feature selection**	**Recognition algorithm**
				Filter	**Processing**	**Detection**			
(Brock, Ohgi, & Lee, 2017)	Ski jumping: nine motion style errors in flight and landing (5 errors during aerial phase/ 4 error during landing phase)	Three: ski jump athletes	85 measured jump motions		1) removal of internal noise 2) sensor alignment to bone direction of mounted segment using standardised calibration measurement 3) neutralisation 4) segmentation of motion streams into jump phases 5) all sensor streams down-sampled by factor of 2 along temporal domain		CNN model - transformed every pre-processed data segment into a multi-channel motion image of size [R, C, D] with D = 3		CNN, SVM
(Buckley et al., 2017)	Running: classification of running form as a non-fatigued or fatigued state	21: 11 females, 10 males, recreationally active	584 extracted stride repetitions labelled as 292 non-fatigued and 292 fatigued	Low-pass Butterworth filter with a frequency cut-off of 5 Hz od order n = 5	Additional signals computed: Euler, pitch, roll, yaw and Quaternion W, X, Y, Z using algorithms on board the Shimmer IMUs. Stride segmentation by an adaptive algorithm		16 time-domain and frequency-domain features computed to describe the 16 IMU signals over each stride repetition.	Wilcoxon Rank Sum Test, the top 20 signal features extracted	RF, SVM, kNN, NB
(Buthe, Blanke, Capkevics, & Tröster, 2016)	Tennis: forehand topspin, forehand slice, backhand topspin, backhand slice, smash, shot steps, side steps	Four: male athletes, three intermediate and 1 advanced	Shots n = 200 Steps n = 640		Shots: discretize data using kMeans algorithm Steps: deadreckoning technique				Shots: LCS Steps: SVM

Table 5 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset sample No.**	**Data pre-processing**			**Feature extraction**	**Feature selection**	**Recognition algorithm**
				Filter	**Processing**	**Detection**			
(Connaghan et al., 2011)	Tennis: serve, forehand, backhand	Eight: two novices, three intermediate, three advanced athletes	2543			Compute length 3D acceleration vector with a W s window around largest absolute magnitude			NB
(Groh, Kautz, & Schuldhaus, 2015)	Skateboarding: ollie, nollie, kickflip, heelflip, pop shove-it, 360-flip	Seven: male, advanced skateboarders as three regular and four goofy stance directions	210		Rider stance correction: x-axes and z-axes for all goofy rider stance data inverted	Accelerometer signal segmented into window lengths 1 s with 0.5 s overlap. Energy of window calculated as sum of squares of all axes. Threshold-based detection defined	Total 54 features calculated: mean, variance, skewness, kurtosis, dominant frequency, bandwidth, x-y-correlation, x-z-correlation, y-z-correlation	Embedded Classification Software Toolbox using the best-first forward selection method	NB, PART, SVM (radial bases kernel), kNN
(Groh, Fleckenstein, & Eskofier, 2016)	Snowboarding: two trick categories (Grinds and Airs) with three trick classes each category	*Part A* Four: male snowboarders, as two regular and two goofy stance directions. *Part B* Seven: male snowboarders, as four regular and three goofy stance directions	275 tricks total (119 Grinds and 156 Airs)		Calibration of accelerometer and gyroscope data using static measurements and rotations about all axes. Rider stance correction: x-axes and z-axes of all goofy rider stance data inverted	Peak detected in accelerometer signal landing after trick. L^1-norm $S\alpha, t$ computed for all times t. Window-based threshold of length 50 samples (0.25s), overlap 49 samples. Threshold determined by LOOCV	Trick category: defined threshold approaches from magnetometer signals Trick class: nine gyroscope signal features of total rotation, rotation for first half of trick, and rotation from s half of trick for each axis		Trick category: NB Trick class: NB, kNN, SVM, C4.5

Table 5 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset sample No.**	**Data pre-processing**			**Feature extraction**	**Feature selection**	**Recognition algorithm**
				Filter	**Processing**	**Detection**			
(Groh, Fleckenstein, Kautz, & Eskofier, 2017)	Skateboarding: 11 trick types, trick fail, resting period	11: skateboard athletes	905 trick events		Calibration. Signal y-axes and z-axes inverted	Accelerometer peaks and gyroscope landing impact signals	Accelerometer: x–z-axes correlation after a landing impact Gyroscope: correlation of the x–y-, x–z- and y–z-axes, and specified rotation features	Trick event interval defined as 1 s before and 0.5 s after landing impact	NB, RF, LSVM, SVM (radial-basis kernel), kNN
(Jensen et al., 2015)	Golf: putt phases, putt event, no-putt event	15: inexperienced golfers	272		Sensor data calibration using the 9DOF Calibration Software (version 2.3). Sensor data transformation using a Direction Cosine Matrix	HMM with sliding windows (500 samples, 1.95 s) with a 50% overlap	31 kinematic parameters from 6D IMU data: (1) phase length and ratios of phase lengths (2) angles and ratios of angles (3) velocity at impact (4) summed acceleration around impact (5) velocity and acceleration profiles in fore-swing		AB
(Jensen, Blank, Kugler, & Eskofier, 2016)	Swimming: rest period, turn, butterfly, backstroke, breaststroke, freestyle	11: high level junior swimmers			Sliding windows between 1 s to 3.5 s with 0.5 s increments. Feature normalization		48D feature vectors per window, computed on each axis: signal energy, min, max, mean, STD, kurtosis, skewness, variance	Best First Search wrapper algorithm	AB, LR, PART, SVM

Table 5 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset sample No.**	**Data pre-processing**			**Feature extraction**	**Feature selection**	**Recognition algorithm**
				Filter	**Processing**	**Detection**			
(Jensen, Prade, & Eskofier, 2013)	Swimming: butterfly, backstroke, breaststroke, freestyle, turns	12: five females and 7 males, high-level swimmers				Spatial energy and head position	48 features total (8 features x 6 axes): mean, STD, variance, energy, kurtosis, skewness, min, max		DT
(Jiao, Wu, Bie, Umek, & Kos, 2018)	Golf: nine swing types	Four: amateur to professional ranked golfers	213 raw samples, 917 samples after augmentation		Dataset augmented to balance swing counts in each class				Vanilla CNN
(Kautz et al., 2017) Machine learning approach	Volleyball: nine shot skill types, one null class	30: 11 females and 19 males, novice to professional	4284	High-pass Butterworth filter with an 8 Hz cut-off frequency	$L1$-norm of the high-passed signal was computed. Signal was smoothed using a low-pass Butterworth filter with a 3 Hz cut-off frequency	Threshold based approach with calculated indicators. C4.5 with LOOCV	39 features: median, mean, STD, skewness, kurtosis, dominant frequency, amplitude of spectrum at dominant frequency, max, min, position of the max, position of the minimum, energy. Pearson correlation coefficients for the correlations between x-axis and y-axis, between x-axis and z-axis, and between y-axis and z-axis	Filter based on the Adjusted Rand Index	SVM, (radial basis kernel function), kNN, Gaussian NB, CART, RF, VOTE

Table 5 continued.

Reference	Sport: target movement(s)	Participants Number: gender, level	Dataset sample No.	Data pre-processing			Feature extraction	Feature selection	Recognition algorithm
				Filter	Processing	Detection			
(Kautz et al., 2017) Deep learning approach	Volleyball: nine shot skill types, one null class	30: 11 females and 19 males, novice to professional	4284		Resampling of raw data				Deep CNN defined as two conv layers with ReLUs and max-pooling, followed by two FC layers with soft-max
(Kelly, Coughlan, Green, & Caulfield, 2012)	Rugby Union: tackle and non-tackle impacts	Nine: professional athletes		Low-pass filter on magnitude signals		Local maxima with an amplitude cut-off of 0.25 Hz	Static window features: max, min, mean, variance, kurtosis, skewness Impact region features: calculated from a window with dynamically calculated start and end points. Impact region signal features: temporal changes in each accelerometer raw data signals		SVM, HCRF, Learning Grid approach with model fusion by AB

Table 5 continued.

Reference	Sport: target movement(s)	Participants Number: gender, level	Dataset sample No.	Data pre-processing			Feature extraction	Feature selection	Recognition algorithm
				Filter	Processing	Detection			
(Kobsar, Osis, Hettinga, & Ferber, 2014)	Running: motion patterns to predict training background and experience level	14, soccer athletes. 16, first time marathon runners. 12, experienced marathon runners	Per participant: 15 s accelerometer data equating to ~20 – 25 footfalls		RMS of accelerations in the vertical, medio-lateral, anteroposterior, and resultant direction calculated. The economy of accelerations determined as the RMS in each axis divided by the gait speed. Outliers adjusted using a Winsorizing technique. All variables standardized to a mean of 0 and a STD of 1		DWT procedure of 5-level wavelet decomposition using Daubechies 5-mother wavelet	PCA	LDA (binary classification)
(Kos & Kramberger, 2017)	Tennis: forehand, backhand, serve	Seven: junior to senior athletes	446			Defined threshold based on two-point derivative of acceleration curves			Unsupervised discriminative analysis
(Ó Conaire et al., 2010)	Tennis: serve, backhand, forehand	Five: elite nationally ranked	300		Normalization of stroke data by rescaling for variance to equal 1	1 s window over accelerometer peaks detected from a threshold approach	Normalized signal x, y, z vectors		SVM (radial basis function kernel), kNN
(O'Reilly et al., 2015)	Squat: correct or incorrect technique and specific technique deviations	22: 4 females and 18 males, with prior experience and regular squat training in regime	682	Low-pass Butterworth filter with a frequency cut-off of 20 Hz			30 features: min and max range accelerometer and gyroscope x, y, z signals, pitch, roll, yaw		Back-propagation NN

Table 5 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset sample No.**	**Data pre-processing**			**Feature extraction**	**Feature selection**	**Recognition algorithm**
				Filter	**Processing**	**Detection**			
(O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017a)	Lunge: discriminate between different levels of lunge performance and identify aberrant techniques	80: 23 females, 57 males, with prior experience and regular lunge training in regime	3440	Low-pass Butterworth filter with frequency cut-off of 20 Hz of order n = 8	3D orientation of IMU computed from all axes using a gradient descent algorithm. Acceleration and gyroscope magnitude calculated. Each exercise repetition resampled to length of 250 samples.		240 features per IMU calculated and extracted including: signal peak, valley, range, mean, standard deviation, skewness, kurtosis, signal energy, level crossing rate, variance, 25th and 75th percentile, median, variance of both the approximate and detailed wavelet coefficients using the Daubechies 5 mother wavelet to level 6		RF
(O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017b)	Deadlifting: technique deviations	135: 41 females and 94 males, with prior lifting experience	2245	Low-pass Butterworth filter with a frequency cut-off of 20 Hz	Rotation quaternions were converted to pitch, roll and yaw signals. Magnitude of acceleration and rotational velocity computed. Time-normalization by exercise repetitions resampled to a length of 250 samples		17 time and frequency domain features each signal: mean, RMS, STD, kurtosis, median, skewness, range, variance, max, min, energy, 25th percentile, 75th percentile, fractal dimension, level crossing-rate, variance of approximate and detailed wavelet coefficients		RF

Table 5 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset sample No.**	**Data pre-processing**			**Feature extraction**	**Feature selection**	**Recognition algorithm**
				Filter	**Processing**	**Detection**			
(Pernek, Kurillo, Stiglic, & Bajcsy, 2015)	Weightlifting: six dumbbell lifting exercises	11: three females and 8 males	~ 2904		Temporal alignment. Uniform resampling of sample rate to 25 Hz		Min, max, range, arithmetic mean, STD, RMS, correlation	Sliding window approach	SVM (Gaussian radial basis function kernel)
(Qaisar et al., 2013)	Cricket: correct and incorrect medium paced bowls	One: medium paced cricket bowler	40		Calibration by filter using signal processing techniques and interpolated to smooth out the filtered data		Mean, mode, STD, peak to peak value, min, max, first deviation, second deviation	K-means clustering	K-means clustering, Markov Model, HMM.
(Rassem, El-Beltagy, & Saleh, 2017)	Cross-country skiing: gears variations	NR	416,737		Data segmented into training, validation, testing set applied with a window size 1 sec with 50% overlap				Recurrent LSTM, CNN, MLP
(Rindal, Seeberg, Tjønnås, Haugnes, & Sandbakk, 2018)	Cross-country skiing: eight technique sub-classes	10: 9 male, 1 female, trained amateurs to professional world-cup skiers	8616	Chest accelerometer data filtered with Gaussian low-pass filter 0.0875 s (1.75 samples) standard deviation in the time domain			Samples were decimated or interpolated into 30 samples per cycle and then appended into one feature vector of 94 samples		NN with three hidden layers of 50, 10, 20 neurons in each layer respectively

Table 5 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset sample No.**	**Data pre-processing**			**Feature extraction**	**Feature selection**	**Recognition algorithm**
				Filter	**Processing**	**Detection**			
(Salman, Qaisar, & Qamar, 2017)	Cricket: detect legal or illegal bowls	14: male cricketers, medium and fast paced bowlers	150	Calibration and filter	Outliers removed using IQR method. Missing values in each attribute replaced with corresponding mean values of attribute, conditional of 10% limit of missing values per attribute before discarded	Data divided into tagged windows corresponding to phases of bowling action. Ball release point was the maxima to denote start process of windowing and tagging	Seven features per axis of accelerometer and gyroscope signals: mean, median, STD, skewness, kurtosis, min, max	Correlation-based feature selection with Greedy search method resulting in the top 21 features	SVM (redial basis function kernel), kNN, NB, RF, NN (three-layer feed-forward)
(Schuldhaus et al., 2015)	Soccer: shot, pass, event leg, support leg, other soccer events	23: male athletes	64 passes, 12 shots	High-pass Butterworth filter		Accelerometer peak detection using a Signal Magnitude Vector. Segmented windows of 1 s around peaks	Four features from each accelerometer axis: mean, variance, skewness, kurtosis		SVM (linear kernel), CART, NB
(Srivastava et al., 2015)	Tennis: forehand, backhand, serve, sub-shot types (flat, topspin, slice)	14: five professional and nine novices	~1000 shots from professional athletes, ~1800 shots from novice athletes			Pan Tomkin's algorithm to isolate shot signal from noise. Accelerometer x-axis differentiated and squared. Moving window integration with window size 3* the sampling rate. Identified potential shot impact region using thresholding			Two Level hierarchical classifier: (1) DTW, (2) QDTW

Table 5 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset sample No.**	**Data pre-processing**			**Feature extraction**	**Feature selection**	**Recognition algorithm**
				Filter	**Processing**	**Detection**			
(Whiteside, Cant, Connolly, & Reid, 2017)	Tennis: serve, forehand (rally, slice, volley), backhand (rally, slice, volley), smash, false shot	19: 8 females and 11 males, junior national development athletes	Per athlete: mean 1504 ± 971		Saturated signals reconstructed using a linear interpolation method. Signals smoothed with 50-point (0.1 sec) moving average.	Threshold algorithm with a window size 0.5 s either side of the detected shot. Shot instances temporally aligned with exported coded vision file.	40 features (5 features across 8 waveforms): min, med, integral, discrete value at time of impact		SVM (linear, quadratic, cubic, Gaussian kernels), CT (10, 25, 50 splits), kNN (k of 1, 3, 5), NN, RF, DA (linear and quadratic)

3D three dimensions, *AB* Adaptive Boosting, *C4.5* decision tree analysis type, *CART* classification and regression tree, *CNN* convolutional neural network, *CT* classification tree, *DA* discriminative analysis, *DOF* degrees of freedom, *DT* decision tree, *DWT* dynamic time warp, *FC* fully-connected, *HCRF* hidden conditional random field, *HMM* Hidden Markov Model, *HZ* hertz, *IMU* inertial measurement unit, *IQR* interquartile range, *kNN* k-Nearest Neighbour, *LCS* Longest Common Subsequence algorithm, *LDA* linear discriminative analysis, *LOOCV* leave-one-out-cross-validation, *LR* logistic regression, *LSTM* long short term memory, *LSVM* linear support vector machine, *MLPs* multi-layer perceptrons, *NB* Naïve Bayesian, *NN* neural network, *NR* not reported, *PART* partial decision tree, *QDTW* Quaternions based Dynamic Time Warping, *ReLUs* rectifier linear unit, *RF* random forests, *RMS* root mean square, *STD* standard deviation, *SVM* Support Vector Machine, *VOTE* vote classifier.

Table 6 Inertial measurement unit study model performance evaluation characteristics.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Anand, Sharma, Srivastava, Kaligounder, & Prakash, 2017)	Detection: precision, recall, F1-score Classification: CA		Detection of squash: • Precision 0.95 • Recall 0.96 • F1- score 0.96 CA: • Tennis: CNN 93.8% • Badminton: BLSTM 78.9% • Squash: BLSTM 94.6%	In-house developed tool to align recorded vision and sensor data to tag shot types in which tagged data serves as ground truth for analysis	
(Adelsberger & Tröster, 2013)	Detection accuracy, CA	75% / 25% train-test dataset split	Detection accuracy: • 100% (when athletes did not move between reps) Classification: • CA 94.117% (between expert and beginner level) Classification: • CA 93.395% (individual thruster instances)	Video footage with performances labelled by a certified coaching expert	Dataset split details: Tennis: training set ~4500 shots by 15 players testing set ~5000 shots by 16 players Badminton: training set ~3500 shots by 20 players testing set ~2000 shots by 14 players Squash: training set ~500 shots by 3 players testing set ~100 shots by 2 players
(Brock & Ohgi, 2017)	Precision, recall, CA, error rate		SVM: CA 52% - 82%	Video control data	For each classifier algorithm, 72 experiments were conducted varying in factor sampling rate (4 variations), windows size (6 variations) and feature selection strategy (3 variations). Error rate defined as the difference between classification accuracy and 1.0
(Brock, Ohgi, & Lee, 2017)	CA, cross-entropy loss	8-fold cross validation	CNN 1 layer: CA 93 ± 0.08%	Jump style annotated by qualified judge under the judging guidelines of the International Skiing Federation	

Table 6 continued.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Buckley et al., 2017)	CA, sensitivity, specificity, F1-score,	LOO-CV 10-K-fold cross validation	Global Classifier: • LIMU lumbar spine CA 75% • IMU right shank CA 70% • IMU left shank CA 67% Personalised classifier: • IMU lumbar spine CA 89% • IMU right shank CA 99% • IMU left shank CA 100%	Manual labelling	Personalised classifiers appear more computationally efficient than global classifiers as they require less training data and memory storage.
(Buthe, Blanke, Capkevics, & Tröster, 2016)	Detection accuracy, confusion matrix, recall, precision, user-specific dataset comparison for train and test	LOO-CV	Step detection accuracy: • Overall 76% • Side steps 96% • Shot steps 63% LOOCV: • Precision 0.49 ± 0.04% • Recall 0.49 ± 0.22% User-specific: • Precision 98% • Recall 87%		Gyroscope signals showed to be more suitable than accelerometer signals to separate shot movements and identify fast foot movements
(Connaghan et al., 2011)	Detection accuracy, CA	10-fold cross validation	Detection accuracy: • Candidate strokes 85% • Non-candidate strokes 85% Classification accuracy: • 3 sensor fusion overall accuracy 90% • Accelerometer 7 player model 97% • Gyroscope 7 player model 76% • Magnetometer 7 player model 76%		Accelerometer signals were the most effective at classifying different skill levels
(Groh, Kautz, & Schuldhaus, 2015)	Detection: sensitivity, specificity Classification: CA, computational effort	LOSO-CV	Detection: • Sensitivity 94.2% • Specificity 99.9% Classification: • CA 97.8% (NB and SVM) Computation effort (lowest): • NB (operations 360, time 6.2 s) • PART (operations 41, time 10.6 s)	Video footage and expert analysis of trick quality	Computational effort defined as the time and required operations for one model run without grid search

Table 6 continued.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Groh, Fleckenstein, & Eskofier, 2016)	Precision, recall, CA	LOSO-CV	Event detection: • Recall 0.99 • Precision 0.368 Trick category classification: • Grind recall 0.966 • Grind precision 0.885 • Airs recall 0.974 • Airs precision 0.910 Trick class CA: • Grind 90.3% (SVM) • Airs 93.3% (kNN)	Video footage	
(Groh, Fleckenstein, Kautz, & Eskofier, 2017)	Detection: precision, recall Classification: CA, confusion matrix	Classification: LOSO-CV	Detection: • Precision 0.669 • Recall 0.964 Classification: • Correct trick execution CA 89.1% (SVM) • All tricks modelled 79.8% CA (RF)	Video footage with manual annotation	
(Jensen et al., 2015)	Detection accuracy, false positive rate		Overall detection rate 68.2%. False positive rate 2.4%	Video footage	Detection rate: $DR = \frac{N_d}{N_p}$ False positive rate: $FPR = \frac{N_m}{N_m + N_p}$ N_d number of detected putts N_p number of performed putts N_m number of misdetected putts
(Jensen, Blank, Kugler, & Eskofier, 2016)	CA	LOSO-CV	Maximum CA 86.5% (SVM) Average CA 82.4% (SVM)	Video footage manually labelled	72 methodological experiments were conducted. A sampling rate of 10.25 Hz and increased window sizes produced higher classification accuracy.
(Jensen, Prade, & Eskofier, 2013)	CA	LOSO-CV	Turn CA 99.8%. Swim stroke CA 95%		

Table 6 continued.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Jiao, Wu, Bie, Umek, & Kos, 2018)	CA, precision, recall	10-fold cross validation	CA 95% Precision 0.95 average Recall 0.95 average F1-score 0.95 average		
(Kautz et al., 2017) Machine learning approach	Confusion matrix, sample accuracy, balanced accuracy, computational time	Detection: LOSO-CV Classification: leave-three-subjects-out cross validation	Sample accuracy 67.2% (VOTE) Balanced accuracy 60.3% (VOTE) Training computational time: • 18.1 ms (NB with feature selection) Class prediction computational time: • 0.53 µs (CART)	Video footage manually labelled	Sample accuracy: $\lambda_s = \frac{\sum_{c=1}^{M} r_c}{\sum_{c=1}^{M} N_c}$ Balanced accuracy: $\lambda_b = \frac{1}{M}\sum_{c=1}^{M}\frac{r_c}{N_c}$ N_c number of samples from class c r_c number of sample from class c classified correctly M number of classes
(Kautz et al., 2017) Deep learning approach	Sample accuracy, balanced accuracy	Leave-two-out cross-validation	Sample accuracy 83.2% Balanced accuracy 79.5%	Video footage manually labelled	
(Kelly, Coughlan, Green, & Caulfield, 2012)	Recall, precision, TP, TN, FP, FN		Learning Grid approach: • Recall 0.933 • Precision 0.958	Video footage manually labelled by the medical staff of the elite rugby union team involved	
(Kobsar, Osis, Hettinga, & Ferber, 2014)	CA	LOO-CV	Training background CA 96.2% Experience level CA 96.4%		

Table 6 continued.

Reference	**Evaluation**	**Cross validation or dataset split approach**	**Performance**	**Ground truth**	**Special remarks**
(Kos & Kramberger, 2017)	CA		Serve CA 98.8%, forehand CA93.5%, backhand CA 98.6%	Video footage	Gyroscope signals were found to be more discriminative between stroke types
(Ó Conaire et al., 2010)	Detection accuracy, CA	LOO-CV	Detection accuracy: 100% Classification: • Right arm data CA 89.41% (kNN) • Full-body data CA 93.44% (kNN)		Data fusion of accelerometer and vision data improved CA: • Vision back viewpoint with full body accelerometer 100% CA (kNN) Data fusion overcame viewpoint sensitivity • Vision trained on side viewpoint and tested on back viewpoint fused with full body accelerometer data 96.71% CA (kNN)
(O'Reilly et al., 2015)	CA, sensitivity, specificity	LOSO-CV	Binary classification: • Sensitivity 64.41% • Specificity 88.01% • CA 80.45% Multi-label classification; • Sensitivity 59.65% • Specificity 94.84% • CA 56.55%	Chartered Physiotherapist evaluation based on the National Strength and Conditioning Association guidelines	
(O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017a)	CA, sensitivity, specificity, out-of-bag error	LOSO-CV	Classify acceptable and aberrant technique Five lower limb IMU set-up: • CA 90% • Sensitivity 80% • Specificity 92% Classify specific technique deviations Five lower limb IMU set-up: • CA 70% • Sensitivity 70% • Specificity 97%	Chartered physiotherapist and strength and conditioning trained practitioner. Correct technique described by the National Strength and Conditioning Association (NSCA) guidelines.	
(O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017b)	CA, sensitivity, specificity	LOSO-CV	Natural technique deviations binary CA: • Global classifier 73% (RF) • Personalized classifier 84% (RF) Natural technique deviations multi-class CA: • Global classifier 54% (RF) • Personalized classifier 78% (RF)	Video footage labelled by a Chartered Physiotherapist	Personalized classifiers outperformed the global classifiers and were more computationally efficient. kNN, SVM, NB tested during analysis against RF, but did not improve results and some caused increased computational times in some cases.

Table 6 continued.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Pernek, Kurillo, Stiglic, & Bajcsy, 2015)	CA, prediction error, confusion matrix	LOSO-CV, 10-fold cross-validation, 75%/ 25%train-test dataset split	Methodology experiments: • CA range 84.2 ± 11.3% to 93.6 ± 0.5% Intensity error: • range 1.2% to 6.6 ± 2.5%	Video footage with manual annotation	A 2 s window size with 50% overlap data processing yielded the best performance results.
(Qaisar et al., 2013)	CA		Overall CA: 90.2% (HMM) • Wrist sensor data 100% • Elbow sensor data 88.24% • Upper arm sensor data 82.35%	Video footage	
(Rassem, El-Beltagy, & Saleh, 2017)	Average testing classification error over the model run. MLP model used as performance benchmark for DL models		Standard LSTM: 1.6% class error value CNN: 2.4% class error value		Data was divided into training, validation and testing sets with a segmentation process applied of window size one second with a 50% overlap.
(Rindal, Seeberg, Tjønnås, Haugnes, & Sandbakk, 2018)	CA, sensitivity, precision, confusion matrix	Validation dataset was used to evaluate which of the 20 trained neural networks to use for final model. Test set created from six different athlete data	CA 99.8% on training dataset CA 96.5% on validation dataset CA 93.9% on combined tests sets	Manual video labelling	Artificially expanded training dataset by taking every cycle in the original training data and created a new cycle by keeping the x-axis and z-axis, whereas the y-axis was flipped resulting in 8616 cycles from the original 4308 training cycles.
(Salman, Qaisar, & Qamar, 2017)	Detection accuracy, CA, recall, precision, F1-score	LOSO-CV	Detection of ball release point 100% accuracy. CA 81 ± 3.12% (SVM) Recall 0.80 (SVM) Precision 0.82 (SVM) F1-score 0.81 (SVM)	Video footage evaluated by an expert cricketer	

Table 6 continued.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Schuldhaus et al., 2015)	CA	LOSO-CV	Set protocol conditions CA (SVM): • Leg type 99.9% • Other events 96.7% • Pass or shot 88.6% Match conditions CA (SVM): • Shot 86.7% • Pass 81.7%	Video footage manually labelled	
(Srivastava et al., 2015)	Detection accuracy, CA		Shot detection accuracy: • Professional 99.58% • Novice 98.96% • Total 99.41% Shot CA: • Class professional player 99.6% • Class novice player 99.3% • Sub-shot types professional player 90.7% • Sub-shot types novice player 86.2%		
(Whiteside, Cant, Connolly, & Reid, 2017)	CA, confusion matrix, precision, recall	10-fold cross-validation	Mean CA (SVM – cubic kernel): • Condition one 97.43 ± 0.24% • Condition two 93.21 ± 0.45%	Video footage manually labelled by a performance analyst	SVM algorithms were constructed using linear, quadratic, cubic and Gaussian kernels, and a one-versus-one approach. kNN classifiers were built using a k of 1,3 and 5. CT were constructed using a maximum of 10, 25 and 50 splits. NN included a conventional single-layer model and multi-layer deep network

CA classification accuracy, *CART* classification and regression tree, *CT* classification tree, *FN* false negative, *FP* false positive, *Hz* hertz, *kNN* k-Nearest Neighbour, *LOO-CV* leave-one-out cross validation, *LOSO-CV* leave-one-subject-out cross validation, *MLP* multi-layer perceptrons, *NB* Naïve Bayesian, *PART* partial decision tree, *RF* random forests, *SVM* Support Vector Machine, *TN* true negative, *TP* true positive, *VOTE* vote classifier.

Table 7 Vision-based study description and model characteristics.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset samples**	**Pre-processing**	**Processing**	**Feature extraction and selection**	**Recognition**
(Bertasius, Park, Yu, & Shi, 2017)	Basketball: some-body shooting a ball, camera wearer possessing the ball, camera wearer shooting the ball	48: male US College players	10.3 hours of recorded vision			Gaussian mixture function	CNN, Multi-path convolutional LSTM
(Couceiro, Dias, Mendes, & Araújo, 2013)	Golf Putting: athlete signature features	Six: male, expert level	180 trial shots (30 trials per athlete)		Darwinian particle swarm optimization method		LDA, QDA, NB with Gaussian distribution, NB with kernel smoothing density estimate, LS-SVM with RBF kernel
(Díaz-Pereira, Gómez-Conde, Escalona, & Olivieri, 2014)	Gymnastics: 10 actions grouped into three categories of jumps, rotations, pre-acrobatics	Eight: junior gymnasts	560 video shots (5 - 7 actions per gymnast)	Motion Vector Flow Instance		PCA and LDA	kNN
(Hachaj, Ogiela, & Koptyra, 2015)	Oyama Karate: 10 classes of actions grouped into 4 defence types, 3 kick types, 3 stands	Six: advanced Oyama karate martial artists	1236	Pre-classification: data pre-processed based on z-scores calculations for each feature value	Segmentation: GDL classifier approach training with an unsupervised R-GDL algorithm. A Baum-Welch algorithm to estimate HMM parameters	Angle-based features	Continuous Gaussian density forward-only HMM classifiers

Table 7 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset samples**	**Pre-processing**	**Processing**	**Feature extraction and selection**	**Recognition**
(Horton, Gudmundsson, Chawla, & Estephan, 2014)	Soccer: Pass quality	Dataset: English Premiership 2007/2008 season games	2932 passes across four matches			Features: basic geometric prediction variables, sequential predictor variables, physiological predictor variables, strategic predictor variables	Multinomial logistic regression, SVM, RUSBoost algorithm
(Ibrahim, Muralidharan, Deng, Vahdat, & Mori, 2016)	Volleyball: six team activity classes, seven individual athlete actions	Dataset: 15 YouTube volleyball videos	1525 annotated frames			CNN	CNN, LSTM
(Kapela, Świetlicka, Rybarczyk, Kolanowski, & O'Connor, 2015)	Rugby, Basketball, Soccer, Cricket, Gaelic football, Hurling: 8 scene types	Dataset	50 hours	Video de-coding: storage of every 5th frame in the buffer		FFT	DT, Feed-forward MLP NN, Elman NN
(Karpathy et al., 2014)	Sports-1M dataset	Dataset	1 million YouTube videos containing 487 classes with 1000 -3000 videos per class	Optimization: Downspur Stochastic Gradient Descent	Data augmentation: (1) crop centre region and resize to 200 x 200 pixels, randomly sampling 170 x 170 region, and randomly flipping images horizontally with 50% probability. (2) subtract constant value of 117 from raw pixel values		CNN (several approaches to fusing data across temporal domains)

Table 7 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset samples**	**Pre-processing**	**Processing**	**Feature extraction and selection**	**Recognition**
(Kasiri-Bidhendi, Fookes, Morgan, Martin, & Sridharan, 2015)	Boxing: 6 punch types of straight, hook, uppercut from both rear and lead hand	Eight: elite orthodox boxers	192 punches (32 for each type)		Detection of body parts: fuzzy inference method based on 2D chamfer distance and geodesic distances	Spatial-temporal features of each punch	RF, Linear SVM, Hierarchical SVM
(Kasiri, Fookes, Sridharan, & Morgan, 2017)	Boxing: 6 punch types of straight, hook, uppercut from both rear and lead hand	14: elite orthodox and southpaw boxers across different weight classes	605 punches		Detection of body parts: fuzzy inference method based on 2D chamfer distance, depth values and geodesic distances	Transition-invariant trajectory features of hand and arm descriptors extracted. Feature ranking for feature reduction experimented using PCA, RF, SVM-reclusive feature eliminator	Multi-class SVM, RF
(Liao, Liao, & Liu, 2003)	Swimming: backstroke, breaststroke, butterfly, freestyle	Dataset	50 clips	Associated limb region detection: RGB images converted to HSV space. Associated skin colour detection: pixels labelled between 0.3 to 1.5 hue values.	Upper body sections isolated using heuristic, threshold approach	LR analysis	DT
(Li et al., 2018)	Golf: key swing gesture detection		Golf front angle swing vision from 553 players, Golf side angle swing vision from 790 players, Baseball swing vision from 3363 players			Multi-scale aggregate channel feature method	AD-DWTAdaBoost Linear SVM

Table 7 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset samples**	**Pre-processing**	**Processing**	**Feature extraction and selection**	**Recognition**
(Lu, Okuma, & Little, 2009)	Ice Hockey: skating movement directions of down, up, left, right	Male unspecified athletes	5609 images of 32 x 32 grayscale images	Tracking: HSV, HOG combined with SVM. Template updating: SPPCA	Multi-target tracking by incorporated SPPCA with an action recognizer using an AB algorithm		SMLR
(Montoliu, Martín-Félez, Torres-Sospedra, & Martínez-Usó, 2015)	Soccer: team activities of ball possessions, quick attack, set pieces	Private dataset: professional Spanish soccer team	Two matches of 90 min each	All camera images combined via algorithmic approach for a unique image covering field length		Bag-of-Words Optical Flow	kNN, SVM, MLP
(Nibali, He, Morgan, & Greenwood, 2017)	Diving: 5 dive properties or rotation type, pose type, number of somersaults, number of twists, handstand beginning inclusion	Dataset: high-level divers from the Australian Institute of Sport	Training set: 25 hours with 4716 non-overlapping dives. Test set: day's footage of 612 dives	Temporal action localisation: TALNN - built from volumetric Convolutional layers. Smoothing: Hann Window Function	Spatial Localisation: full regression, partial regression, segmentation, and Global constraints (RANSAC algorithm).		C3D volumetric convolutional network (3x3x3 kernels, ReLUs, dropouts)
(Ó Conaire et al., 2010)	Tennis: serve, forehand, backhand	Five: elite nationally ranked			Contour features: back-ground subtraction and image morphology	Player foreground region divided into 16 pie segments centred on player centroid and normalization	SVM with RBF kernel, kNN
(Ramanathan et al., 2015)	Basketball: 11 match activity classes and frame key player detection	Dataset: 257 NCAA games from YouTube	1143 training clips, 856 validation clips, 2256 testing clips	Each clip subsampled to six fps at four seconds in length		Each video-frame represented by a 1024-dimensional feature vector. Appearance features extracted using the Inception7 (Szegedy & Ibarz, 2015) network and spatially pooling th e response from the lower layer. Features corresponded to a 32x32 spatial histogram combined with a spatial pyramid	LSTM and BLSTM RNNs

Table 7 continued.

Reference	**Sport: target movement(s)**	**Participants** Number: gender, level	**Dataset samples**	**Pre-processing**	**Processing**	**Feature extraction and selection**	**Recognition**
(Reily, Zhang, & Hoff, 2017)	Gymnastics: Pommel horse routine spinning	Unspecified male gymnasts	10115 frames recorded as 16-bit PNG images, organized into 39 routines	DOI segmentation: (1) Parazen window (2) Identified signal peaks padded with neighbourhood 10% max depth		SAD3D: The gymnast in each frame is described by features: (1) width of their silhouette, (2) height of their silhouette, (3–4) depth values at the leftmost and rightmost ends of the silhouette, (5– 8) shift in the left-most x, right-most x, upper y, and lower y coordinates compared to the previous frame.	SVM with radial basis function kernel. Smoothing techniques after classification
(Shah, Chokalingam, Paluri, & Pradeep, 2007)	Tennis: forehand, backhand, other	Dataset: male and female unspecified athletes	150 games each clipped to 10 min segments	Optical flow calculated between consecutive frames	Image segmentation and weight calculation by global adaptive thresholding. Player appearance modelling by Expectation Maximization algorithm	Oriented histogram of skeletonized binary images of athletes	SVM with RBF kernel
(Tora, Chen, & Little, 2017)	Ice Hockey: dump in, dump out, pass, shot, loose puck recovery	Dataset: National Hockey League videos	2507 training events, 250 testing events			Features extracted by the fc7 layers of AlexNet (Krizhevsky, Sutskever, & Hinton, 2012). Max-pooling of features of individual players in frames to incorporate player interactions	LSTM

Table 7 continued.

Reference	Sport: target movement(s)	Participants Number: gender, level	Dataset samples	Pre-processing	Processing	Feature extraction and selection	Recognition
(Victor, He, Morgan, & Miniutti, 2017)	Swimming: backstroke, breaststroke, butterfly, freestyle Tennis: stroke detection	Datasets: Swimming: 40 athletes Tennis: 4 athletes	15k swimming strokes labelled in 650k fames. 1.3k tennis strokes labelled in 270 frames	Swimming: pre-processed using Hough transform as in (Sha, Lucey, Morgan, Pease, & Sridharan, 2013) to extract the lanes from colour information. Tennis: excluded unlabelled tennis strokes from input dataset. Input data frames down sampled to 192 x 128 pixels	Model parameters initialized. Adedelta optimizer. MSE loss function. All frame's pixels encoded in YUV colour-space and down sampled to 128 x 48		Regression: CNN with a base architecture based off the VGG-B CNN (Simonyan & Zisserman, 2014)
(Yao & Fei-Fei, 2010)	Human-object interaction sport activities: cricket defensive shot, cricket bowling, croquet shot, tennis forehand, tennis serve, volleyball smash	Dataset	350 images (50 images per 6 classes)	Gaussian over the number of edges and randomization of initialization connectivity to different starting points	Hill-climbing approach with a Tabu list	Parameter estimation with a max-margin learning method	Composition inference method
(Zhu, Xu, Gao, & Huang, 2006)	Tennis: left and right swings	Professional tennis athletes	6035 frames of 1099 left swing strokes and 1071 right swing strokes		Player tracking: SVR particle filter and background subtraction.	Motion descriptor extraction: optical flow computed using Horn-Sckunck algorithm with half-wave rectification and Gaussian smoothing. Feature discrimination: slice-based optical flow histograms	SVM

2D two dimensional, *BLSTM* bidirectional LSTM, *CNN* convolutional neural network, *DOI* Depth of interest segmentation, *DT* decision tree, *ELU* Exponential Linear Units, *FFT* Fast Fourier Transform, *GDL* Gesture Description Language, *HMM* Hidden Markov Model, *HOG* Histogram of Oriented Gradients, *HSV* Hue-Saturation-Value-Colour-Histogram, *kNN* k-Nearest Neighbour, *LDA* linear discriminative analysis, *LR* logistic regression, *LS-SVM* least squares support vector machine, *MLP* multi-layer perceptron, *NB* Naïve Bayesian, *NN* neural network, *PCA* principal component analysis, *PNG* Portable Network Graphics, *QDA* quadratic discriminative analysis, *RBF* radial basis function, *RF* random forests, *RUSBoost* Random Under Sampling Boosting, *SAD3D* Silhouette Activity Descriptor in 3 Dimensions, *SPPCA* Switching Probabilistic Principal Component Analysis, *SVM* Support Vector Machine, *SVR* Support Vector Regression.

Table 8 Vision-based study model performance evaluation characteristics.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Bertasius, Park, Yu, & Shi, 2017)	F1-score	24 videos for training dataset, 24 videos for testing dataset	Basketball event detection mean F1-score 0.625. Basketball athlete performance evaluation model F1-score 0.793.	Manual labelling and athlete performance assessment by a former professional basketball player	Compared model's performance to first-person activity recognition baselines and a video activity recognition baseline C3D
(Couceiro, Dias, Mendes, & Araújo, 2013)	Confusion matrix, ROC		LS-SVM overall best performance		1) five classifiers evaluated for detecting signature patterns 2) best classifier method applied to extract individual golf putt signatures
(Díaz-Pereira, Gómez-Conde, Escalona, & Olivieri, 2014)	True/ false recognition rates for binary classification, sensitivity, specificity	10-fold cross validation	Specificity 85% overall Sensitivity 90% overall		
(Hachaj, Ogiela, & Koptyra, 2015)	CA, confusion matrix	LOO-CV	Overall CA range across classes 93 ± 7% to 100% (four-state HMM)		Five HMM classifiers tested with number of hidden states ranging from 1 (GMM) to 5
(Horton, Gudmundsson, Chawla, & Estephan, 2014)	CA, precision, recall, F1-score	80%/ 20% train-test dataset split. Tests set was stratified so per class frequency was consistent with the distribution in training examples	Three-class model 85.5% (SVM)	Labelled data of pass events. Rating of pass quality by observers (6-point Likert Scale) Cohen's Kappa for heuristic measure of agreement between ratings	Experiments conducted using two labelling schemes: 1) six-class labels assigned by observers. 2) three-class scheme (aggregation of six-classes) Test dataset was stratified so per-class frequency consistent with distribution in training dataset.
(Ibrahim, Muralidharan, Deng, Vahdat, & Mori, 2016)	CA, confusion matrix	2/3rd of total data as training set, 1/3rd as testing set	51.1% CA		Compared model performance to several baseline models

Table 8 continued.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Kapela, Świetlicka, Rybarczyk, Kolanowski, & O'Connor, 2015)	Modified accuracy (focused around detection performance), precision, modified precision		Overall precision 0.96	Manual annotation	$\text{Modified accuracy} = \frac{(\text{DE} - \text{DTE})}{\text{NE}}$ $\text{Precision} = \frac{\text{DTE}}{\text{DE}}$ $\text{Modified precision} = \frac{\text{DTE}}{\text{NE}}$
Karpathey et al. (Karpathy et al., 2014)	Prediction classification accuracy %, per-class average precision, confusion matrix	Dataset split: 70% training set, 10% validation set, 20% test set	CNN model average CA 63.9% Slow fusion model CA 60.9%	Labelled data classes	
(Kasiri-Bidhendi, Fookes, Morgan, Martin, & Sridharan, 2015)	CA, confusion matrix	LOO-CV Model trained on data from seven participants and tested on withheld data from one participant	Hierarchal SVM CA 92 – 96%	Start and end frames of each punch labelled by expert analysts	
(Kasiri, Fookes, Sridharan, & Morgan, 2017)	CA, feature numbers, confusion matrix		Hierarchical SVM CA 97.3%	Start and end frames of each punch labelled by expert analysts	
(Liao, Liao, & Liu, 2003)	Developed scoring system based on measure of proximity to the prominent feature of a specific style				
(Li et al., 2018)	CA, precision, recall, computational time	Cross-validation (not specified). Dataset split: 80% train/ 10% validation/ 10% test set	CA 97% Average recognition time of 2.38 ms		

Table 8 continued.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Lu, Okuma, & Little, 2009)	CA, average computing speed, confusion matrix		SMLR and HOG approach CA 76.37% Computing speed: average total time classification image 0.206s (SMLR and HOG approach)	Manual image retrieval and division into the four classes	Compared developed model against benchmark action recognizers.
(Montoliu, Martín-Félez, Torres-Sospedra, & Martínez-Usó, 2015)	CA	5-fold cross-validation, LOO-CV	RF CA 92.89 ± 0.2%	Manual vision annotation by expert	
(Nibali, He, Morgan, & Greenwood, 2017)	CA, precision, recall, F1-score		Dive property CA from 86.89 - 100%	Labelled training data	Segmentation works best (spatial localisation). Dilated convolutions boosted CA.
(Ó Conaire et al., 2010)	CA	LOO-CV	Back viewpoint CA 98.67% (kNN) Side viewpoint CA 95% (kNN)		Data fusion of accelerometer and vision data improved CA: • Vision back viewpoint with full body accelerometer CA 100% (kNN) Data fusion overcame viewpoint sensitivity • Vision trained on side viewpoint and tested on back viewpoint fused with full body accelerometer data CA 96.71% (kNN)
(Ramanathan et al., 2015)	Mean average precision	Hyperparameters chosen by cross-validating on the validation dataset	Event classification 0.516 mean average precision Event detection 0.435 mean average precision Key player attention 0.618 mean average precision	Manually labelled videos through an Amazon Mechanical Turk task	Event classification from isolated video clips was compared against different control setting and baseline models

Table 8 continued.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Reily, Zhang, & Hoff, 2017)	CA, computational time, error rates (RMSE, average absolute), approach tested on CAD60 dataset benchmark		ID depth interest CA 97.8% Spin detection CA 93.81% Smoothing processing improved spin CA to 94.83%. Spin consistency performance analysis in comparison to ground truth RMSE 12.9942 ms from ground truth timestamp.	Manually labelled dataset	Study model reduces late stage data amount processing to perform calculations on 37.8% of the original data.
(Shah, Chokalingam, Paluri, & Pradeep, 2007)	CA, confusion matrix		Forehand CA 97.24% Backhand CA 96.42% No stroke CA 98.02%	Manually labelled segment frames	Model computational performance speed was 20 fps
(Tora, Chen, & Little, 2017)	CA, Confusion matrix		Overall 49.2% CA		Model compared to several baseline models
(Victor, He, Morgan, & Miniutti, 2017)	F1-score, average frame distance, average distance to smoothed	80%/ 20% train-test dataset split	Swimming F1-score 0.922 Tennis F1-score 0.977	Manually labelled dataset by expert analysts	Experimented with how temporal information incorporated into the model, data input style, and three smoothing functions. Developed model tested and validated on tennis stroke dataset
(Yao & Fei-Fei, 2010)	CA, compared developed model to previous published benchmarks and a baseline measure (bag-of-words with a linear SVM)	60%/ 40% train-test dataset split	Activity CA 83.3%	Labelled training dataset	

Table 8 continued.

Reference	Evaluation	Cross validation or dataset split approach	Performance	Ground truth	Special remarks
(Zhu, Xu, Gao, & Huang, 2006)	Precision, recall		Tennis stroke classification using video frames: • Left recall 84.08%, • Left precision 89.80% • Right recall 90.20%, • Right precision 84.66%. Tennis stroke classification using action clips: • Left recall 87.50%, • Left precision 90.74% • Right recall 89.80%, • Right precision 86.27%		
CA classification accuracy, *CNN* convolutional neural network, *DE* detected events, *DTE* detected true events, *GMM* Gaussian mixture model, *HMM* Hidden Markov Model, *kNN* k-Nearest Neighbour, *LOO-CV* leave-one-out cross validation, *LOSO-CV* leave-one-subject-out cross validation, *LS-SVM* least squares support vector machine, *NE* number of events, *RF* random forests, *ROC* receiver operation characteristic curve, *SVM* Support Vector Machine.					

Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance

Running title:

Machine and deep learning for sport movement recognition review

Abstract

Objective assessment of an athlete's performance is of importance in elite sports to facilitate detailed analysis. The implementation of automated detection and recognition of sport-specific movements overcomes the limitations associated with manual performance analysis methods. The object of this study was to systematically review the literature on machine and deep learning for sport-specific movement recognition using inertial measurement unit (IMU) and, or computer vision data inputs. A search of multiple databases was undertaken. Included studies must have investigated a sport-specific movement and analysed via machine or deep learning methods for model development. A total of 52 studies met the inclusion and exclusion criteria. Data pre-processing, processing, model development and evaluation methods varied across the studies. Model development for movement recognition were predominantly undertaken using supervised classification approaches. A kernel form of the Support Vector Machine algorithm was used in 53% of IMU and 50% of vision-based studies. Twelve studies used a deep learning method as a form of Convolutional Neural Network algorithm and one study also adopted a Long Short Term Memory architecture in their model. The adaptation of experimental set-up, data pre-processing, and model development methods are best considered in relation to the characteristics of the targeted sports movement(s).

Key Words:

Sport movement classification; inertial sensors; computer vision; machine learning; performance analysis.

1. Introduction

Performance analysis in sport science has experienced considerable recent changes, due largely to access to improved technology and increased applications from computer science. Manual notational analysis or coding in sports, even when performed by trained analysts, has limitations. Such methods are typically time intensive, subjective in nature, and prone to human error and bias. Automating sport movement recognition and its application towards coding has the potential to enhance both the efficiency and accuracy of sport performance analysis. The potential automation of recognising human movements, commonly referred to as human activity recognition (HAR), can be achieved through machine or deep learning model approaches. Common data inputs are obtained from inertial measurement units (IMUs) or vision. Detection refers to the identification of a targeted instance, i.e., tennis strokes within a continuous data input signal (Bulling, Blanke, & Schiele, 2014). Recognition or classification of movements involves further interpretations and labelled predictions of the identified instance (Bulling et al., 2014; Bux, Angelov, & Habib, 2017), i.e., differentiating tennis strokes as a forehand or backhand. In machine and deep learning, a model represents the statistical operations involved in the development of an automated prediction task (LeCun, Yoshua, & Geoffrey, 2015; Shalev-Shwartz & Ben-David, 2014).

Human activities detected by inertial sensing devices and computer vision are represented as wave signal features corresponding to specific actions, which can be logged and extracted. Human movement activities are considered hierarchically structured and can be broken down to basic movements. Therefore, the context of signal use, intra-class variability, and inter-class similarity between activities require consideration during experimental set-up and model development. Wearable IMUs contain a combination of accelerometer, gyroscope, and magnetometer sensors measuring along one to three axes. These sensors quantify acceleration, angular velocity, and the direction and orientation of travel respectively (Gastin, McLean, Breed, & Spittle, 2014). These sensors can capture repeated movement patterns during sport training and competitions (Camomilla, Bergamini, Fantozzi, & Vannozzi, 2018; Chambers, Gabbett, Cole, & Beard, 2015; J. F. Wagner, 2018). Advantages include being wireless, lightweight and self-contained in operation. Inertial measurement units have been utilised in quantifying physical output and tackling impacts in Australian Rules football (Gastin et al., 2014; Gastin, McLean, Spittle, & Breed, 2013) and rugby

(Gabbett, Jenkins, & Abernethy, 2012, 2011; Howe, Aughey, Hopkins, Stewart, & Cavanagh, 2017; Hulin, Gabbett, Johnston, & Jenkins, 2017). Other applications include swimming analysis (Mooney, Corley, Godfrey, Quinlan, & ÓLaighin, 2015), golf swing kinematics (Lai, Hetchl, Wei, Ball, & McLaughlin, 2011), over-ground running speeds (Wixted, Billing, & James, 2010), full motions in alpine skiing (Yu et al., 2016); and the detection and evaluation of cricket bowling (McNamara, Gabbett, Blanch, & Kelly, 2017; McNamara, Gabbett, Chapman, Naughton, & Farhart, 2015; Wixted, Portus, Spratford, & James, 2011).

Computer vision has applications for performance analysis including player tracking, semantic analysis, and movement analysis (Stein et al., 2018; Thomas, Gade, Moeslund, Carr, & Hilton, 2017). Automated movement recognition approaches require several pre-processing steps including athlete detection and tracking, temporal cropping and targeted action recognition, which are dependent upon the sport and footage type (Barris & Button, 2008; Saba & Altameem, 2013; Thomas et al., 2017). Several challenges including occlusion, viewpoint variations, and environmental conditions may impact results, depending on the camera set-up (Poppe, 2010; Zhang et al., 2017). Developing models to automate sports-vision coding may improve resource efficiency and reduce feedback times. For example, coaches and athletes involved in time-intensive notational tasks, including post-swim race analysis, may benefit from rapid objective feedback before the next race in the event program (Liao, Liao, & Liu, 2003; Victor, He, Morgan, & Miniutti, 2017). For detecting and recognising movements, body worn sensor signals do not suffer from the same environmental constraints and stationary set-up of video cameras. Furthermore, multiple sensors located on different body segments have been argued to provide more specific signal representations of targeted movements (J. B. Yang, Nguyen, San, Li, & Shonali, 2015). But it is not clear if this is solely conclusive, and the use of body worn sensors in some sport competitions may be impractical or not possible.

Machine learning algorithms learn from data input for automated model building and perform tasks without being explicitly programmed. The algorithm goal is to output a response function $h\sigma(\bar{x})$ that will predict a ground truth variable y from an input vector of variables $\bar{x}$. Models are run for classification techniques to predict a target class (Kotsiantis, Zaharakis, & Pintelas, 2007), or regression to predict discrete or continuous values. Models are aimed at finding an optimal set of parameters σ to describe the response function, and then make predictions on unseen unlabelled data input. Within these, model training approaches can generally run as supervised learning, unsupervised learning or semi-supervised learning (Mohammed, Khan, & Bashier, 2016; Sze, Chen, Yang, & Emer, 2017).

Processing raw data is limited for conventional machine learning algorithms, as they are unable to effectively be trained on abstract and high-dimensional data that is inconsistent, contains missing values or noisy artefacts (Bux et al., 2017; Kautz, 2017). Consequently, several pre-processing stages are required to create a suitable data form for input into the classifier algorithm (Figo, Diniz, Ferreira, & Cardoso, 2010). Filtering (Figo et al., 2010; Wundersitz, Gastin, Robertson, Davey, & Netto, 2015), window capture durations (Mitchell, Monaghan, & O'Connor, 2013; Preece, Goulermas, Kenney, & Howard, 2009; Wundersitz, Josman, et al., 2015), and signal frequency cut-offs (Wundersitz, Gastin, Richter, Robertson, & Netto, 2015; Wundersitz, Gastin, Robertson, et al., 2015) are common techniques applied prior to data prior to dynamic human movement recognition. Well-established filters for processing motion signal data include the Kalman filter (Kautz, 2017; Titterton & Weston, 2009; D. Wagner, Kalischewski, Velten, & Kummert, 2017) and a Fourier transform filter (Preece, Goulermas, Kenney, Howard, et al., 2009) such as a fast Fourier transform (Kapela, Świetlicka, Rybarczyk, Kolanowski, & O'Connor, 2015; Preece, Goulermas, Kenney, & Howard, 2009). Near real-time processing benefits from reducing memory requirements, computational demands, and essential bandwidth during whole model implementation. Signal feature extraction and selection favours faster processing by reducing the signals to the critical features that can discriminate the targeted activities (Bulling et al., 2014). Feature extraction involves identifying the key features that help maximise classifier success, and removing features that have minimal impact in the model (Mannini & Sabatini, 2010). Thus, feature selection involves constructing data representations in subspaces with reduced dimensions. These identified variables are represented in a compact feature variable (Mannini & Sabatini, 2010). Common methods include principal component analysis (PCA) (Gløersen, Myklebust, Hallén, & Federolf, 2018; Young & Reinkensmeyer, 2014), vector coding techniques (Hafer & Boyer, 2017) and empirical cumulative distribution functions (ECDF) (Plötz, Hammerla, & Olivier, 2011). An ECDF approach has been shown to be advantageous over PCA as it derives representations of raw input independent of the absolute data ranges, whereas PCA is known to have reduced performance when the input data is not properly normalised (Plötz et al., 2011). For further detailed information on the acquisition, filtering and analysis of IMU data for sports application and vision-based human activity recognition, see (Kautz, 2017) and (Bux et al., 2017), respectively.

Deep learning is a division of machine learning, characterised by deeper neural network model architectures and are inspired by the biological neural networks of the human brain (Bengio, 2013; LeCun et al., 2015; Sze et al., 2017). The deeper hierarchical models create a profound architecture of multiple hidden layers based on representative learning with several processing and abstraction layers (Bux et al., 2017; J. B. Yang et al., 2015). These computational models allow data input features to be automatically extracted from raw data and transformed to handle unstructured data, including vision (LeCun et al., 2015; Ravi, Wong, Lo, & Yang, 2016). This direct input avoids several processing steps required in machine learning during training and testing, therefore reducing overall computational times. A current key element within deep learning is backpropagation (Hecht-Nielsen, 1989; LeCun, Bottou, Orr, & Müller, 1998). Backpropagation is a fast and computationally efficient algorithm, using gradient descent, that allows training deep neural networks to be tractable (Sze et al., 2017). Human activity recognition has mainly been performed using conventional machine learning classifiers. Recently, deep learning techniques have enhanced the bench mark and applications for IMUs (Kautz et al., 2017; Ravi et al., 2016; Ronao & Cho, 2016; J. B. Yang et al., 2015; Zebin, Scully, & Ozanyan, 2016; Zeng et al., 2014) and vision (Ji, Yang, Yu, & Xu, 2013; Karpathy et al., 2014a; Krizhevsky, Sutskever, & Hinton, 2012; Nibali, He, Morgan, & Greenwood, 2017) in human movement recognition producing more superior model performance accuracy.

The objective of this study was to systematically review the literature investigating sport-specific automated movement detection and recognition. The review focusses on the various technologies, analysis techniques and performance outcome measures utilised. There are several reviews within this field that are sensor-based including wearable IMUs for lower limb biomechanics and exercises (Fong & Chan, 2010; M. O'Reilly, Caulfield, Ward, Johnston, & Doherty, 2018), swimming analysis (Magalhaes, Vannozzi, Gatta, & Fantozzi, 2015; Mooney et al., 2015), quantifying sporting movements (Chambers et al., 2015) and physical activity monitoring (C. C. Yang & Hsu, 2010). A recent systematic review has provided an evaluation on the in-field use of inertial-based sensors for various performance evaluation applications (Camomilla et al., 2018). Vision-based methods for human activity recognition (Aggarwal & Xia, 2014; Bux et al., 2017; Ke et al., 2013; Zhang et al., 2017), semantic human activity recognition (Ziaeefard & Bergevin, 2015) and motion analysis in sport (Barris & Button, 2008) have also been reviewed. However, to date, there is no systematic review across sport-specific movement detection and recognition via machine or deep learning. Specifically, incorporating IMUs and vision-based data input, focussing on in-field applications as opposed to laboratory-based protocols and detailing the analysis and machine learning methods used.

Considering the growth in research and potential field applications, such a review is required to understand the research area. This review aims to characterise the evolving techniques and inform researchers of possible improvements in sports analysis applications. Specifically: 1) What is the current scope for IMUs and computer vision in sport movement detection and recognition? 2) Which methodologies, inclusive of signal processing and model learning techniques, have been used to achieve sport movement recognition? 3) Which evaluation methods have been used in assessing the performance of these developed models?

2. Methods

2.1 Search strategy

The preferred PRISMA recommendations (Moher, Liberati, Tetzlaff, Altman, & Group, 2009) for systematic reviews were used. A literature search was undertaken by the first author on the following databases; IEEE Xplore, PubMed, ScienceDirect, Scopus, Academic Search Premier, and Computer and Applied Science Complete. The searched terms were categorised in order to define the specific participants, methodology and evaluated outcome measure in-line with the review aims. Searches used a combination of key words with AND/OR phrases which are detailed in Table 1. Searches were filtered for studies from January 2000 to May 2018 as no relevant studies were identified prior to this. Further studies were manually identified from the bibliographies of database-searched studies identified from the abstract screen phase, known as snowballing. Table 2 provides the inclusion and exclusion criteria of this review.

*****Table 1 near here: Key word search term strings per database *****

*****Table 2 near here: Inclusion and exclusion criteria*****

2.2 Data extraction

The first author extracted and collated the relevant information from the full manuscripts identified for final review. A total of 18 parameters were extracted from the 52 research studies, including the title, author, year of publication, sport, participant details, sport movement target(s), device specifications, device sample frequency, pre-processing methods, processing methods, feature selected, feature extraction, machine learning model used, model evaluation, model performance accuracy, validation method, samples collected, and computational information. A customised Microsoft Excel^{TM} spreadsheet was developed to categorise the relevant extracted information from each study. Participant characteristics of number of participants, gender, and competition level, then if applicable a further descriptor specific to a sport, for example, ‘medium-paced cricket bowler’. Athlete and participant experience level was categorised as written in the corresponding study to avoid misrepresentations. The age of participants was not considered an important characteristic required for model development. The individual ability in which the movement is performed accounts for the discriminative signal features associated with the movements. For the purposes of this review, a sport-specific movement was defined from a team or individual sport, and training activities associated with a particular sport. For example, weight-lifting as strength training, recognised under the Global Association of International Sports Federations. The targeted sports and specific movements were defined for either detection or recognition. Model development techniques used included pre-processing methods to transform data to a more suitable form for analysis, processing stages to segment data for identified target activities, feature extraction and selections techniques, and the learning algorithm(s). Model evaluation measures extracted were the model performance assessment techniques used, ground-truth validation comparison, number of data samples collected, and the model performance outcomes results reported. If studies ran multiple experiments using several algorithms, only the superior algorithm and relevant results were reported as the best method. This was done so in the interest of concise reporting to highlight favourable method approaches (Sprager & Juric, 2015). Any further relevant results or information identified from the studies was included as a special remark (Sprager & Juric, 2015). Hardware and specification information extracted included the IMU or video equipment used, number of units, attachment of sensors (IMUs), sample frequency, and sensor data types used in analysis (IMUs). Studies identified and full data extracted were reviewed by a second author.

3. Results

An outline of the search results and study exclusions has been provided in Fig 1. Of the initial database search which identified 4885 results, a final 52 studies met criteria for inclusion in this review. Of these, 29 used IMUs and 22 were vision-based. One study (Ó Conaire et al., 2010) used both sensors and vision for model development separately then together via data fusion. Tables 3 - 8 provide a description of the characteristics of the reviewed studies, detailed in the following sections.

***** Fig 1 near here: PRISMA flow diagram *****

3.1 Experimental design

A variety of sports and their associated sport-specific movements were investigated, implementing various experimental designs as presented in Tables 5 and 7. Across the studies, sports reported were tennis (n = 10), cricket (n = 3), weightlifting or strength training (n = 6), swimming (n = 4), skateboarding (n = 2), ski jumping (n = 2), snowboarding (n = 1), golf (n = 4), volleyball (n = 2), rugby (n = 2), ice hockey (n = 2), gymnastics (n = 2), karate (n = 1), basketball (n = 3), Gaelic football (n = 1), hurling (n = 1), boxing (n = 2), running (n = 2), diving (n = 1), squash (n = 1), badminton (n = 1), cross-country skiing (n = 2) and soccer (n = 4). The Sports 1-M dataset (Karpathy et al., 2014b) was also reported, which consists of 1,133,158 video URLs annotated automatically with 487 sport labels using the YouTube Topic API. A dominant approach was the classification of main characterising actions for each sport. For example, serve, forehand, backhand strokes in tennis (Connaghan et al., 2011; Kos & Kramberger, 2017; Ó Conaire et al., 2010; Shah, Chokalingam, Paluri, & Pradeep, 2007; Srivastava et al., 2015), and the four competition strokes in swimming (Jensen, Blank, Kugler, & Eskofier, 2016; Jensen, Prade, & Eskofier, 2013; Liao et al., 2003; Victor et al., 2017). Several studies further classified sub-categories of actions. For example, three further classes of the two main classified snowboarding trick types Grinds and Airs (Groh, Fleckenstein, & Eskofier, 2016), and further classifying the main tennis stroke types as either flat, topspin or slice

(Srivastava et al., 2015). Semantic descriptors were reported for classification models that predicted athlete training background, experience and fatigue level. These included running (Buckley et al., 2017; Kobsar, Osis, Hettinga, & Ferber, 2014), rating of gymnastic routines (Reily, Zhang, & Hoff, 2017), soccer pass classification based on its quality (Horton, Gudmundsson, Chawla, & Estephan, 2014), cricket bowling legality (Qaisar et al., 2013; Salman, Qaisar, & Qamar, 2017), ski jump error analysis (Brock & Ohgi, 2017; Brock, Ohgi, & Lee, 2017) and strength training technique deviations (M. A. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017a; M. O'Reilly et al., 2015; M. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017). One approach (Yao & Fei-Fei, 2010), encoded the mutual context of human pose and sporting equipment using semantics, to facilitate the detection and classification of movements including a cricket bat and batsman coupled movements.

Total participant numbers for IMU-based studies ranged from one (Qaisar et al., 2013) to 30 (Kautz et al., 2017). Reported data individual instance sample sizes for sensor studies ranged from 150 (Salman et al., 2017) to 416, 737 (Rassem, El-Beltagy, & Saleh, 2017). Vision-based studies that explicitly reported total participant details ranged from five (Ó Conaire et al., 2010) to 40 (Victor et al., 2017). Vision dataset sample sizes varied across studies, from 50 individual action clips (Liao et al., 2003) to 15, 000 (Victor et al., 2017). One study (Karpathy et al., 2014a) used the publicly available Sports-1M, as previously described. Vision-based studies also reported datasets in total time, 10.3 hours (Bertasius, Park, Yu, & Shi, 2017), 3 hours (Montoliu, Martín-Félez, Torres-Sospedra, & Martínez-Usó, 2015), 1, 500 minutes (Shah et al., 2007), and 50 hours (Kapela et al., 2015), and by frame numbers, 6, 035 frames (Zhu, Xu, Gao, & Huang, 2006) and 10, 115 frames (Reily et al., 2017).

3.2 Inertial measurement unit specifications

A range of commercially available and custom-built IMUs were used in the IMU-based studies (n= 30), as presented in Table 3. Of these, 23% reported using a custom-built sensor. Of the IMU-based studies, the number of sensors mounted or attached to each participant or sporting equipment piece ranged from one to nine. The majority of studies (n= 22) provided adequate details of sensor specifications including sensor type, axes, measurement range, and sample rate used. At least one characteristic of sensor measurement range or sample rate used in data collection was missing from eight studies. All studies used triaxial sensors and collected accelerometer data. For analysis and model development, individual sensor data consisted of only accelerometer data (n = 8), both accelerometer and gyroscope data (n = 15), and accelerometer, gyroscope and magnetometer data (n = 7). The individual sensor measurement ranges reported for accelerometer were ± 1.5 g to ± 16 g, gyroscope ± 500 ◦/s to ± 2000 ◦/s, magnetometer ± 1200 µT or 1.2 to 4 Ga. Individual sensor sample rates ranged from 10 Hz to 1000 Hz for accelerometers, 10 Hz to 500 Hz for gyroscopes and 50 Hz to 500 Hz for magnetometers.

***** Table 3 near here*****

3.3 Vision capture specification

Several experimental set-ups and specifications were reported in the total 23 vision-based studies (Table 4). Modality was predominately red, green, blue (RGB) cameras. Depth cameras were utilised (Kasiri-Bidhendi, Fookes, Morgan, Martin, & Sridharan, 2015; Kasiri, Fookes, Sridharan, & Morgan, 2017; Reily et al., 2017), which add depth perception for 3-dimensional image mapping. Seven studies clearly reported the use of a single camera set-up (Couceiro, Dias, Mendes, & Araújo, 2013; Díaz-Pereira, Gómez-Conde, Escalona, & Olivieri, 2014; Hachaj, Ogiela, & Koptyra, 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Nibali et al., 2017; Reily et al., 2017). One study reported 16 stationary positioned cameras at a ‘bird’s eye view’ (Montoliu et al., 2015), and Ó Conaire et al. (2010) reported the use of one overhead and 8 stationary cameras around a tennis court baseline, although data from two cameras were only used in final analysis due to occlusion issues. Sample frequency and, or pixel resolution were reported in seven of the studies (Couceiro et al., 2013; Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Montoliu et al., 2015; Victor et al., 2017; Zhu et al., 2006), with sample frequencies ranging from 30 Hz to 210 Hz.

***** Table 4 near here*****

3.4 Inertial measurement unit recognition model development methods

Key stages of model development from data pre-processing to recognition techniques for IMU-based studies are presented in Table 5. Data pre-processing filters were reported as either a low-pass filter (n = 7) (Adelsberger & Tröster, 2013; Buckley et al., 2017; Kelly, Coughlan, Green, & Caulfield, 2012; M. A. O’Reilly et al., 2017a; M. O’Reilly et al., 2015, 2017; Rindal, Seeberg, Tjønnås, Haugnes, & Sandbakk, 2018), high-pass filter (n = 2) (Kautz et al., 2017; Schuldhaus et al., 2015), or calibration with a filter (Salman et al., 2017). Processing methods were reported in 67% of the IMU-based studies (Adelsberger & Tröster, 2013; Anand, Sharma, Srivastava, Kaligounder, & Prakash, 2017; Brock et al., 2017; Buckley et al., 2017; Buthe, Blanke, Capkevics, & Tröster, 2016; Groh et al., 2016; Groh, Fleckenstein, Kautz, & Eskofier, 2017; Groh, Kautz, & Schuldhaus, 2015; Jensen et al., 2016, 2015; Jiao, Wu, Bie, Umek, & Kos, 2018; Kautz et al., 2017; Kobsar et al., 2014; M. A. O’Reilly et al., 2017a; M. O’Reilly et al., 2017; Ó Conaire et al., 2010; Pernek, Kurillo, Stiglic, & Bajcsy, 2015; Qaisar et al., 2013; Salman et al., 2017; Schuldhaus et al., 2015). Methods included, calibration of data (Groh et al., 2016, 2017; Jensen et al., 2015; Qaisar et al., 2013), a one-second window centred around identified activity peaks in the signal (Adelsberger & Tröster, 2013; Schuldhaus et al., 2015), temporal alignment (Pernek et al., 2015), normalisation (Ó Conaire et al., 2010), outlier adjustment (Kobsar et al., 2014) or removal (Salman et al., 2017), and sliding windows ranging from one to 3.5 seconds across the data (Jensen et al., 2016). The three studies that investigated trick classification in skateboarding (Groh et al., 2017, 2015) and snowboarding (Groh et al., 2016) corrected data for different rider board stance styles, termed Regular or Goofy, by inverting signal axes.

Movement detection methods were specifically reported in 16 studies (Adelsberger & Tröster, 2013; Anand et al., 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015, Jensen et al., 2013, 2015; Kautz et al., 2017; Kelly et al., 2012; Kos & Kramberger, 2017; Ó Conaire et al., 2010; Rindal et al., 2018; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside, Cant, Connolly, & Reid, 2017). Detection methods included thresholding (n = 5), windowing segmenting (n = 4), and a combination of threshold and windowing techniques (n = 5).

Signal feature extraction techniques were reported in 80% of the studies, with the number of feature parameters in a vector ranging from a vector of normalised X, Y, Z accelerometer signals (Ó Conaire et al., 2010) to 240 features (M. A. O’Reilly et al., 2017a). Further feature selection to reduce the dimensionality of the feature vector was used in 11 studies. Both feature extraction and selection methods varied considerably across the literature (Table 5).

Algorithms trialled for movement recognition were diverse across the literature (Table 5). Supervised classification using a kernel form of Support Vector Machine (SVM) was most prevalent (n = 16) (Adelsberger & Tröster, 2013; Brock & Ohgi, 2017; Brock et al., 2017; Buckley et al., 2017; Buthe et al., 2016; Groh et al., 2016, 2017, 2015; Jensen et al., 2016; Kautz et al., 2017; Kelly et al., 2012; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside et al., 2017). The next highest tested were Naïve Bayesian (NB) (n = 8) (Buckley et al., 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015; Kautz et al., 2017; Salman et al., 2017; Schuldhaus et al., 2015) and k-Nearest Neighbour (kNN) (n = 8) (Buckley et al., 2017; Groh et al., 2016, 2017, 2015; Kautz et al., 2017; Ó Conaire et al., 2010; Salman et al., 2017; Whiteside et al., 2017), followed by Random Forests (RF) (n = 7) (Buckley et al., 2017; Groh et al., 2017; Kautz et al., 2017; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2017; Salman et al., 2017; Whiteside et al., 2017). Supervised learning algorithms were the most common (n = 29). One study used an unsupervised discriminative analysis approach for detection and classification of tennis strokes (Kos & Kramberger, 2017). Five IMU-based study investigated a deep learning approach including using Convolutional Neural Networks (CNN) (Anand et al., 2017; Brock et al., 2017; Jiao et al., 2018; Kautz et al., 2017; Rassem et al., 2017) and Long Short Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) architectures (Rassem et al., 2017; Sharma, Srivastava, Anand, Prakash, & Kaligounder, 2017). In order to assess the effectiveness of the various classifiers from each study, model performance measures quantify and visualise the predictive performance as reported in the following section.

***** Table 5 near here*****

3.5 Inertial measurement unit recognition model evaluation

Reported performance evaluations of developed models across the IMU-based studies are shown in Table 6. Classification accuracy, as a percentage score for the number of correct predictions by total number of predictions made, was the main model evaluation measure (n = 24). Classification accuracies across studies ranged between 52% (Brock & Ohgi, 2017) to 100% (Buckley et al., 2017). Generally, the reported highest accuracy for a specific movement was ≥ 90% (n = 17) (Adelsberger & Tröster, 2013; Anand et al., 2017; Buckley et al., 2017; Connaghan et al., 2011; Groh et al., 2015; Jensen et al., 2013; Jiao et al., 2018; Kobsar et al., 2014; Kos & Kramberger, 2017; M. A. O'Reilly et al., 2017a; Ó Conaire et al., 2010; Pernek et al., 2015; Qaisar et al., 2013; Rindal et al., 2018;

Schuldhaus et al., 2015; Srivastava et al., 2015; Whiteside et al., 2017) and ≥ 80% to 90% (n = 7) (Brock & Ohgi, 2017; Brock et al., 2017; Groh et al., 2017; Jensen et al., 2016; M. O'Reilly et al., 2015, 2017; Salman et al., 2017). As an estimate of the generalised performance of a trained model on $n - x$ samples, a form of leave-one-out cross validation (LOO-CV) was used in 47% of studies (Buthe et al., 2016; Groh et al., 2016, 2017, 2015, Jensen et al., 2016, 2013; Kobsar et al., 2014; M. O'Reilly et al., 2015, 2017; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015). Precision, specificity and sensitivity (also referred to as recall) evaluations were derived for detection (n = 6) and classification models (n = 10). Visualisation of prediction results in the form of a confusion matrix featured in six studies (Buthe et al., 2016; Groh et al., 2017; Kautz et al., 2017; Pernek et al., 2015; Rindal et al., 2018; Whiteside et al., 2017).

***** Table 6 near here*****

3.6 Vision recognition model development methods

Numerous processing and recognition methods featured across the vision-based studies to transform and isolated relevant input data (Table 7). Pre-processing stages were reported in 14 of studies, and another varied 13 studies also provided details of processing techniques. Signal feature extraction and feature selection methods used were reported in 78% of studies.

Both machine (n = 16) and deep learning (n = 7) algorithms were used to recognise movements from vision data. Of these, a kernel form of the SVM algorithm was most common in the studies (n = 10) (Couceiro et al., 2013; Horton et al., 2014; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Li et al., 2018; Montoliu et al., 2015; M. A. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017b; Ó Conaire et al., 2010; Reily et al., 2017; Shah et al., 2007; Zhu et al., 2006). Other algorithms included kNN (n = 3) (Díaz-Pereira et al., 2014; Montoliu et al., 2015; Ó Conaire et al., 2010), decision tree (DT) (n = 2) (Kapela et al., 2015; Liao et al., 2003), RF (n = 2) (Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017), and Multilayer Perceptron (MLP) (n = 2) (Kapela et al., 2015; Montoliu et al., 2015). Deep learning was investigated in seven studies (Bertasius et al., 2017; Ibrahim, Muralidharan, Deng, Vahdat, & Mori, 2016; Karpathy et al., 2014a; Nibali et al., 2017;

Ramanathan et al., 2015; Tora, Chen, & Little, 2017; Victor et al., 2017) of which used CNNs or LSTM RNNs as the core model structure.

***** Table 7 near here*****

3.7 Vision recognition model evaluation

Performance evaluation methods and results for vision-based studies are reported in Table 8. As with IMU-based studies, classification accuracy was the common method for model evaluations, featured in 61%. Classification accuracies were reported between 60.9% (Karpathy et al., 2014a) and 100% (Hachaj et al., 2015; Nibali et al., 2017). In grouping the reported highest accuracies for a specific movement that were ≥ 90% (n = 9) (Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Li et al., 2018; Montoliu et al., 2015; Nibali et al., 2017; Ó Conaire et al., 2010; Reily et al., 2017; Shah et al., 2007), and ≥ 80% to 90% (n = 2) (Horton et al., 2014; Yao & Fei-Fei, 2010). A confusion matrix as a visualisation of model prediction results was used in nine studies (Couceiro et al., 2013; Hachaj et al., 2015; Ibrahim et al., 2016; Karpathy et al., 2014a; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Lu, Okuma, & Little, 2009; Shah et al., 2007; Tora et al., 2017). Two studies assessed and reported their model computational average speed (Lu et al., 2009) and time (Reily et al., 2017).

***** Table 8 near here*****

4 Discussion

The aim of this systematic review was to evaluate the use of machine and deep learning for sport-specific movement recognition from IMUs and, or computer vision data inputs. Overall, the search yielded 52 studies, categorised as 29 which used IMUs, 22 vision-based and one study using both IMUs and vision. Automation or semi-automated sport movement recognition models working in near-real time is of particular interest to avoid the error, cost and time associated with manual methods. Evident in the literature, models are trending towards the potential to provide optimised objective assessments of athletic movement for technical and tactical evaluations. The majority of studies achieved favourable movement recognition results for the main characterising actions of a sport, with several studies exploring further applications such as an automated skill quality evaluation or judgement scoring, for example automated ski jump error evaluation (Brock et al., 2017).

Experimental set-up of IMU placement and numbers assigned per participant varied between sporting actions. The sensor attachment locations set by researchers appeared dependent upon the specific sporting conditions and movements, presumably to gain optimal signal data. Proper fixation and alignment of the sensor axes with limb anatomical axes is important in reducing signal error (Fong & Chan, 2010). The attachment site hence requires a biomechanical basis for accuracy of the movement being targeted to obtain reliable data. Single or multiple sensor use per person also impacts model development trade-off between accuracy, analysis complexity, and computational speed or demands. In tennis studies, specificity whilst using a single sensor was demonstrated by mounting the IMU on the wrist or forearm of the racquet arm (Connaghan et al., 2011; Kos & Kramberger, 2017; Srivastava et al., 2015; Whiteside et al., 2017). A single sensor may also be mounted in a low-profile manner on sporting equipment (Groh et al., 2016, 2017, 2015; Jensen et al., 2015). Unobtrusive use of a single IMU to capture generalised movements across the whole body was demonstrated, with an IMU mounted on the posterior head in swimming (Jensen et al., 2016, 2013), lower back during running (Kobsar et al., 2014), and between the shoulder blades in rugby union (Kelly et al., 2012).

The majority of vision-based studies opted for a single camera set-up of RGB modality. Data output from a single camera as opposed to multiple minimises the volume of data to process, therefore reducing computational effort. However, detailed features may go uncaptured, particularly in team sport competition which consists of multiple individuals participating in the capture space at one time. In contrast, a multiple camera set-up reduces limitations including occlusion and viewpoint variations. However, this may also increase the complexity of the processing and model computational stages. Therefore, a trade-off between computational demands and movement recording accuracy often needs to be made. As stated earlier, the placement of cameras needs to suit the biomechanical nature of the targeted movement and the environment situated in. Common camera capture systems used in sports science research such as Vicon Nexus (Oxford, UK) and OptiTrack (Oregon, USA) were not present in this review. As this review targeted studies investigating during on-field or in-situation sporting contexts, efficiency in data collection is key for routine applications in training and competition. A simple portable RGB camera is easy to set-up in a dynamic and changing environment, such as different soccer pitches, rather than a multiple capture system such as Vicon that requires calibrated precision and are substantially more expensive.

Data acquisition and type from an IMU during analysis appears to influence model trade-off between accuracy and computational effort of performance. The use of accelerometer, gyroscope or magnetometer data may depend upon the movement properties analysed. Within tennis studies, gyroscope signals were the most efficient at discriminating between stroke types (Buthe et al., 2016; Kos & Kramberger, 2017) and detecting an athlete's fast feet court actions (Buthe et al., 2016). In contrast, accelerometer signals produced higher classification accuracies in classifying tennis stroke skills levels (Connaghan et al., 2011). The authors expected lower gyroscope classification accuracies as temporal orientation measures between skill levels of tennis strokes will differ (Connaghan et al., 2011). Conversely, data fusion from all three individual sensors resulted in a more superior model for classifying advanced, intermediate and novices tennis player strokes (Connaghan et al., 2011). Fusion of accelerometer and vision data also resulted in a higher classification accuracy for tennis stroke recognition (Ó Conaire et al., 2010).

Supervised learning approaches were dominant across IMU and vision-based studies. This is a method which involves a labelled ground truth training dataset typically manually annotated by sport analysts. Labelled data instances were recorded as up to 15, 000 for vision-based (Victor et al., 2017) and 416, 737 for sensor-based (Rassem et al., 2017) studies. Generation of a training data set for supervised learning can be a tedious and labour-intensive task. It is further complicated if multiple sensors or cameras are incorporated for several targeted movements. A semi-supervised or unsupervised learning approach may be advantageous as data labelling is minimal or not required, potentially reducing human errors in annotation. An unsupervised approach could suit specific problems to explain key data features, via clustering (Mohammed et al., 2016; Sze et al., 2017). Results computed by an unsupervised model (Kos, Ženko, Vlaj, & Kramberger, 2016) for tennis serve, forehand and backhand stroke classification compared favourbaly well against a proposed supervised approach (Connaghan et al., 2011).

Recognition of sport-specific movements was primarily achieved using conventional machine learning approaches, however nine studies implemented deep learning algorithms. It is expected that future model developments will progressively feature deep learning approaches due to development of better hardware, and the advantages of more efficient model learning on large data inputs (Sze et al., 2017). Convolutional Neural networks (CNN) (LeCun, Bottou, Bengio, & Haffner, 1998) were the core structure of five of the seven deep learning study models. Briefly, convolution applies several filters, known as kernels, to automatically extract features from raw data inputs. This process works under four key ideas to achieve optimised results: local connection, shared weights, pooling and applying several layers (LeCun et al., 2015; J. B. Yang et al., 2015). Machine learning classifiers modelled with generic hand-crafted features, were compared against a CNN for classifying nine beach volleyball actions using IMUs (Kautz et al., 2017). Unsatisfactory results were obtained from the machine learning model, and the CNN markedly achieved higher classification accuracies (Kautz et al., 2017). The CNN model produced the shortest overall computation times, requiring less computational effort on the same hardware (Kautz et al., 2017). Vision-based CNN models have also shown favourable results when compared to a machine learning study baseline (Karpathy et al., 2014a; Nibali et al., 2017; Victor et al., 2017). Specifically, consistency between a swim stroke detection model for continuous videos in swimming which was then applied to tennis strokes with no domain-specific settings introduced (Victor et al., 2017). The authors of this training approach (Victor et al., 2017) anticipate that this could be applied to train separate models for other sports movement detection as the CNN model demonstrated the ability to learn to process continuous videos into a 1-D signal with the signal peaks corresponding to arbitrary events. General human activity recognition using CNN have shown to be a superior approach over conventional machine learning algorithms using both IMUs (Ravi et al., 2016; J. B. Yang et al., 2015; Zebin et al., 2016; Zeng et al., 2014; Zheng, Liu, Chen, Ge, & Zhao, 2014) and computer vision (Ji et al., 2013; Krizhevsky et al., 2012; LeCun et al., 2015). As machine learning algorithms extract heuristic features requiring domain knowledge, this creates shallower features which can make it harder to infer high-level and context aware activities (J. B. Yang et al., 2015). Given the previously described advantages of deep learning algorithms which apply to CNN, and the recent results of deep learning, future model developments may benefit from exploring these methods in comparison to current bench mark models.

Model performance outcome metrics quantify and visualise the error rate between the predicted outcome and true measure. Comparatively, a kernel form of an SVM was the most common classifier implemented and produced the strongest machine learning approach model prediction accuracies across both IMU (Adelsberger & Tröster, 2013; Brock & Ohgi, 2017; Buthe et al., 2016; Groh et al., 2016, 2017, 2015; Jensen et al., 2016; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside et al., 2017) and vision-based study designs (Horton et al., 2014; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Li et al., 2018; Reily et al., 2017; Shah et al., 2007; Zhu et al., 2006). Classification accuracy was the most common reported measure followed by confusion matrices, as ways to clearly present prediction results and derive further measures of performance. Further measures included sensitivity (also called recall), specificity and precision, whereby results closer to 1.0 indicate superior model performance, compared to 0.0 or poor model performance. The F1-score (also called a F-measure or F-score) conveys the balances between the precision and sensitivity of a model. An in-depth analysis performance metrics specific to human activity recognition is located elsewhere (Minnen, Westeyn, Starner, Ward, & Lukowicz, 2006; Ward, Lukowicz, & Gellersen, 2011). Use of specific evaluation methods depends upon the data type. Conventional performance measures of error rate are generally unsuitable for models developed from skewed training data (Provost & Fawcett, 2001). Using conventional performance measures in this context will only take the default decision threshold on a model trained, if there is an uneven class distribution this may lead to imprecision (Provost & Fawcett, 2001; Seiffert, Khoshgoftaar, Van Hulse, & Napolitano, 2008). Alternative evaluators including Receiver Operating Characteristics (ROC) curves and its single numeric measure, Area Under ROC Curve (AUC), report model performances across all decision thresholds (Seiffert et al., 2008). Making evaluations between study methodology have inherent complications due to each formulating their own experimental parameter settings, feature vectors and training algorithms for movement recognition. The No-Free-Lunch theorems are important deductions in the formation of models for supervised machine learning (David H. Wolpert, 1996), and search and optimisation algorithms (D H Wolpert & Macready, 1997). The theorems broadly reference that there is no ‘one model’ that will perform optimally across all recognition problems. Therefore, experiments with multiple model development methods for a particular problem is recommended. The use of prior knowledge about the task should be implemented to adapt the model input and model parameters in order to improve overall model success (Shalev-Shwartz & Ben-David, 2014).

Acquisition of athlete specific information, including statistics on number, type and intensity of actions, may be of use in the monitoring of athlete load. Other potential applications include personalised movement technique analysis (M. O'Reilly et al., 2017), automated performance evaluation scoring (Reily et al., 2017) and team ball sports pass quality rating (Horton et al., 2014). However, one challenge lies in delivering consistent, individualised models across team field sports that are dynamic in nature. For example, classification of soccer shots and passes showed a decline in model performance accuracy from a closed environment to a dynamic match setting (Schuldhaus et al., 2015). A method to overcome accuracy limitations in dynamic team field sports associated with solely using IMUs or vision may be to implement data fusion (Ó Conaire et al., 2010). Furthermore, vision and deep learning approaches have demonstrated the ability to track and classify team sport collective court activities and individual player specific movements in volleyball (Ibrahim et al., 2016), basketball (Ramanathan et al., 2015) and ice hockey (Tora et al., 2017). Accounting for methods from experimental set-up to model evaluation, previous reported models should be considered and adapted based on the current problem. Furthermore, the balance between model computational efficiency, results accuracy and complexity trade-offs calculations are an important factor.

In the present study, meta-analysis was considered however variability across developed model parameter reporting and evaluation methods did not allow for this to be undertaken. As this field expands and further methodological approaches are investigated, it would be practical to review analysis approaches both within and between sports. This review was delimited to machine and deep learning approaches to sport movement detection and recognition. However, statistical or parametric approaches not considered here such as discriminative functional analysis may also show efficacy for sport-specific movement recognition. However, as the field of machine learning is a rapidly developing area shown to produce superior results, a review encompassing all possible other methods may have complicated the reporting. Since sport-specific movements and their environments alter the data acquisition and analysis, the sports and movements reported in the present study provide an overview of the current field implementations.

5 Conclusions

This systematic review reported on the literature using machine and deep learning methods to automate sport-specific movement recognition. In addressing the research questions, both IMUs and computer vision have demonstrated capacity in improving the information gained from sport movement and skill recognition for performance analysis. A range of methods for model development were used across the reviewed studies producing varying results. Conventional machine learning algorithms such as Support Vector Machines and Neural Networks were most commonly implemented. Yet in those studies which applied deep learning algorithms such as Convolutional Neural Networks, these methods outperformed the machine learning algorithms in comparison. Typically, the models were evaluated using a leave-one-out cross validation method and reported model performances as a classification accuracy score. Intuitively, the adaptation of experimental set-up, data processing, and recognition methods used are best considered in relation to the characteristics of the sport and targeted movement(s). Consulting current models within or similar to the targeted sport and movement is of benefit to address bench mark model performances and identify areas for improvement. The application within the sporting domain of machine learning and automated sport analysis coding for consistent uniform usage appears currently a challenging prospect, considering the dynamic nature, equipment restrictions and varying environments arising in different sports.

Future work may look to adopt, adapt and expand on current models associated with a specific sports movement to work towards flexible models for mainstream analysis implementation. Investigation of deep learning methods in comparison to conventional machine learning algorithms would be of particular interest to evaluate if the trend of superior performances is beneficial for sport-specific movement recognition. Analysis as to whether IMUs and vision alone or together yield enhanced results in relation to a specific sport and its implementation efficiency would also be of value. In consideration of the reported study information, this review can assist future researchers in broadening investigative approaches for sports performance analysis as a potential to enhancing upon current methods.

Acknowledgements

Not applicable.

Disclosure statement

No potential conflict of interest was reported by the authors. However, the last author is co-author of three studies included in this systematic review.

Funding

The authors received no specific funding for this work.

References

Adelsberger, R., & Tröster, G. (2013). Experts lift differently: Classification of weight-lifting athletes. In *2013 IEEE International Conference on Body Sensor Networks* (pp. 1–6). Cambridge, MA: Body Sensor Networks (BSN). https://doi.org/10.1109/BSN.2013.6575458

Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3D data: A review. *Pattern Recognition Letters*, *48*, 70–80. https://doi.org/10.1016/j.patrec.2014.04.011

Anand, A., Sharma, M., Srivastava, R., Kaligounder, L., & Prakash, D. (2017). Wearable motion sensor based analysis of swing sports. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 261–267). https://doi.org/10.1109/ICMLA.2017.0-149

Barris, S., & Button, C. (2008). A review of vision-based motion analysis in sport. *Sports Medicine*, *38*(12), 1025–1043. https://doi.org/10.2165/00007256-200838120-00006

Bengio, Y. (2013). Deep learning of representations: Looking forward. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7978 LNAI*, 1–37. https://doi.org/10.1007/978-3-642-39593-2_1

Bertasius, G., Park, H. S., Yu, S. X., & Shi, J. (2017). Am I a baller? Basketball performance assessment from first-person videos. *Proceedings of the IEEE International Conference on Computer Vision*, 2196–2204. https://doi.org/10.1109/ICCV.2017.239

Brock, H., & Ohgi, Y. (2017). Assessing motion style errors in ski jumping using inertial sensor devices. *IEEE Sensors Journal*, (99), 1–11. https://doi.org/10.1109/JSEN.2017.2699162

Brock, H., Ohgi, Y., & Lee, J. (2017). Learning to judge like a human: convolutional networks for classification of ski jumping errors. *Proceedings of the 2017 ACM International Symposium on Wearable Computers - ISWC '17*, 106–113. https://doi.org/10.1145/3123021.3123038

Buckley, C., O'Reilly, M. A., Whelan, D., Vallely Farrell, A., Clark, L., Longo, V., … Caulfield, B. (2017). Binary classification of running fatigue using a single inertial measurement unit. In *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks* (pp. 197–201). IEEE. https://doi.org/10.1109/BSN.2017.7936040

Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, *46*(3), 1–33. https://doi.org/http://dx.doi.org/10.1145/2499621

Buthe, L., Blanke, U., Capkevics, H., & Tröster, G. (2016). A wearable sensing system for timing analysis in tennis. In *BSN 2016 - 13th Annual Body Sensor Networks Conference* (pp. 43–48). San Francisco, CA. https://doi.org/10.1109/BSN.2016.7516230

Bux, A., Angelov, P., & Habib, Z. (2017). Vision based human activity recognition: A review. In P. Angelov, A. Gegov, C. Jayne, & Q. Shen (Eds.), *Advances in Computational Intelligence Systems: Contributions Presented at the 16th UK Workshop on Computational Intelligence* (pp. 341–371). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-46562-3_23

Camomilla, V., Bergamini, E., Fantozzi, S., & Vannozzi, G. (2018). Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: a systematic review. *Sensors*, *18*(3), 873. https://doi.org/10.3390/s18030873

Chambers, R., Gabbett, T., Cole, M. H., & Beard, A. (2015). The use of wearable microsensors to quantify sport-specific movements. *Sports Medicine*, *45*(7), 1065–1081. https://doi.org/10.1007/s40279-015-0332-9

Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., Walsh, M., & O'Mathuna, C. (2011). Multi-sensor classification of tennis strokes. *Journal of IEEE Sensors*, 1437–1440.

Couceiro, M. S., Dias, G., Mendes, R., & Araújo, D. (2013). Accuracy of pattern detection methods in the performance of golf putting. *Journal of Motor Behavior*, *45*(1), 37–53. https://doi.org/10.1080/00222895.2012.740100

Díaz-Pereira, M. P., Gómez-Conde, I., Escalona, M., & Olivieri, D. N. (2014). Automatic recognition and scoring of olympic rhythmic gymnastic movements. *Human Movement Science*, *34*(1), 63–80. https://doi.org/10.1016/j.humov.2014.01.001

Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. M. P. (2010). Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, *14*(7), 645–662. https://doi.org/10.1007/s00779-010-0293-9

Fong, D. T.-P., & Chan, Y.-Y. (2010). The use of wearable inertial motion sensors in human lower limb biomechanics studies: A systematic review. *Sensors*, *10*(12), 11556–11565. https://doi.org/10.3390/s101211556

Gabbett, T., Jenkins, D., & Abernethy, B. (2012). Physical demands of professional rugby league training and competition using microtechnology. *Journal of Science and Medicine in Sport*, *15*, 80–86. https://doi.org/10.1016/j.jsams.2011.07.004

Gabbett, T., Jenkins, D. G., & Abernethy, B. (2011). Physical collisions and injury in professional rugby league match-play. *Journal of Science and Medicine in Sport*, *14*, 210–215. https://doi.org/10.1016/j.jsams.2011.01.002

Gastin, P. B., McLean, O. C., Breed, R. V., & Spittle, M. (2014). Tackle and impact detection in elite Australian football using wearable microsensor technology. *Journal of Sports Sciences*, *32*(10), 947–953. https://doi.org/10.1080/02640414.2013.868920

Gastin, P. B., McLean, O. C., Spittle, M., & Breed, R. V. (2013). Quantification of tackling demands in professional Australian football using integrated wearable athlete tracking technology. *Journal of Science and Medicine in Sport*, *16*(6), 589–593. https://doi.org/10.1016/j.jsams.2013.01.007

Gløersen, Ø., Myklebust, H., Hallén, J., & Federolf, P. (2018). Technique analysis in elite athletes using principal component analysis. *Journal of Sports Sciences*, *36*(2), 229–237. https://doi.org/10.1080/02640414.2017.1298826

Groh, B. H., Fleckenstein, M., & Eskofier, B. M. (2016). Wearable trick classification in freestyle snowboarding. In *13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)* (pp. 89–93). IEEE. https://doi.org/10.1109/BSN.2016.7516238

Groh, B. H., Fleckenstein, M., Kautz, T., & Eskofier, B. M. (2017). Classification and visualization of skateboard tricks using wearable sensors. *Pervasive and Mobile Computing*, *40*, 42–55. https://doi.org/10.1016/j.pmcj.2017.05.007

Groh, B. H., Kautz, T., & Schuldhaus, D. (2015). IMU-based trick classification in skateboarding. In *KDD Workshop on Large-Scale Sports Analytics*.

Hachaj, T., Ogiela, M. R., & Koptyra, K. (2015). Application of assistive computer vision methods to Oyama karate techniques recognition. *Symmetry*, *7*(4), 1670–1698. https://doi.org/10.3390/sym7041670

Hafer, J. F., & Boyer, K. A. (2017). Variability of segment coordination using a vector coding technique: reliability analysis for treadmill walking and running. *Gait and Posture*, *51*, 222–227. https://doi.org/10.1016/j.gaitpost.2016.11.004

Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. *Proceedings Of The International Joint Conference On Neural Networks*, *1*, 593–605. https://doi.org/10.1109/IJCNN.1989.118638

Hochreiter, S., & Schmidhuber, J. J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1–32. https://doi.org/10.1162/neco.1997.9.8.1735

Horton, M., Gudmundsson, J., Chawla, S., & Estephan, J. (2014). Classification of passes in football matches using spatiotemporal data. *ArXiv Preprint ArXiv:1407.5093*. https://doi.org/10.1145/3105576

Howe, S. T., Aughey, R. J., Hopkins, W. G., Stewart, A. M., & Cavanagh, B. P. (2017). Quantifying important differences in athlete movement during collision-based team sports: Accelerometers outperform global positioning systems. In *2017 IEEE International Symposium on Inertial Sensors and Systems* (pp. 1–4). Kauai, HI, USA: IEEE. https://doi.org/10.1109/ISISS.2017.7935655

Hulin, B. T., Gabbett, T., Johnston, R. D., & Jenkins, D. G. (2017). Wearable microtechnology can accurately identify collision events during professional rugby league match-play. *Journal of Science and Medicine in Sport*, *20*(7), 638–642. https://doi.org/http://dx.doi.org/10.1016/j.jsams.2016.11.006

Ibrahim, M., Muralidharan, S., Deng, Z., Vahdat, A., & Mori, G. (2016). A Hierarchical Deep

Temporal Model for Group Activity Recognition. *Cvpr*, 1971–1980. https://doi.org/10.1109/CVPR.2016.217

Jensen, U., Blank, P., Kugler, P., & Eskofier, B. M. (2016). Unobtrusive and energy-efficient swimming exercise tracking using on-node processing. *IEEE Sensors Journal*, *16*(10), 3972–3980. https://doi.org/10.1109/JSEN.2016.2530019

Jensen, U., Prade, F., & Eskofier, B. M. (2013). Classification of kinematic swimming data with emphasis on resource consumption. In *2013 IEEE International Conference on Body Sensor Networks, BSN 2013*. https://doi.org/10.1109/BSN.2013.6575501

Jensen, U., Schmidt, M., Hennig, M., Dassler, F. A., Jaitner, T., & Eskofier, B. M. (2015). An IMU-based mobile system for golf putt analysis. *Sports Engineering*, *18*(2), 123–133. https://doi.org/10.1007/s12283-015-0171-9

Ji, S., Yang, M., Yu, K., & Xu, W. (2013). 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 221–231. https://doi.org/10.1109/TPAMI.2012.59

Jiao, L., Wu, H., Bie, R., Umek, A., & Kos, A. (2018). Multi-sensor Golf Swing Classification Using Deep CNN. *Procedia Computer Science*, *129*, 59–65. https://doi.org/10.1016/j.procs.2018.03.046

Kapela, R., Świetlicka, A., Rybarczyk, A., Kolanowski, K., & O’Connor, N. E. (2015). Real-time event classification in field sport videos. *Signal Processing: Image Communication*, *35*, 35–45. https://doi.org/10.1016/j.image.2015.04.005

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014a). Large-scale video classification with convolutional neural networks. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference On*, 1725–1732. https://doi.org/10.1109/CVPR.2014.223

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014b). Large-scale video classification with convolutional nural networks. Retrieved December 18, 2017, from http://cs.stanford.edu/people/karpathy/deepvideo/

Kasiri-Bidhendi, S., Fookes, C., Morgan, S., Martin, D. T., & Sridharan, S. (2015). Combat sports analytics: Boxing punch classification using overhead depth imagery. In *2015 IEEE International Conference on Image Processing (ICIP)* (pp. 4545–4549). Quebec City, Canada: IEEE. https://doi.org/10.1109/ICIP.2015.7351667

Kasiri, S., Fookes, C., Sridharan, S., & Morgan, S. (2017). Fine-grained action recognition of boxing punches from depth imagery. *Computer Vision and Image Understanding*, *159*, 143–153. https://doi.org/10.1016/j.cviu.2017.04.007

Kautz, T. (2017). Acquisition, filtering and analysis of positional and inertial data in sports. *FAU Studies in Computer Science*, *2*.

Kautz, T., Groh, B. H., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B. M. (2017). Activity recognition in beach volleyball using a deep convolutional neural network. *Data Mining and Knowledge Discovery*, 1–28. https://doi.org/10.1007/s10618-017-0495-0

Ke, S. R., Thuc, H., Lee, Y. J., Hwang, J. N., Yoo, J. H., & Choi, K. H. (2013). A review on video-based human activity recognition. *Computers*, *2*, 88–131. https://doi.org/10.3390/computers2020088

Kelly, D., Coughlan, G. F., Green, B. S., & Caulfield, B. (2012). Automatic detection of collisions in elite level rugby union using a wearable sensing device. *Sports Engineering*, *15*(2), 81–92. Retrieved from https://0-link-springer-com.library.vu.edu.au/article/10.1007%2Fs12283-012-0088-5

Kobsar, D., Osis, S. T., Hettinga, B. A., & Ferber, R. (2014). Classification accuracy of a single tri-axial accelerometer for training background and experience level in runners. *Journal of Biomechanics*, *47*(10), 2508–2511. https://doi.org/10.1016/j.jbiomech.2014.04.017

Kos, M., & Kramberger, I. (2017). A wearable device and system for movement and biometric data Acquisition for sports applications. *IEEE Access*, 1–1. https://doi.org/10.1109/ACCESS.2017.2675538

Kos, M., Ženko, J., Vlaj, D., & Kramberger, I. (2016). Tennis stroke detection and classification using miniature wearable IMU device. In *International Conference on Systems, Signals, and Image Processing*. https://doi.org/10.1109/IWSSIP.2016.7502764

Kotsiantis, S., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, *31*, 501–520. https://doi.org/10.1115/1.1559160

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances In Neural Information Processing Systems*, 1097–

1105. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

Lai, D. T. H., Hetchl, M., Wei, X., Ball, K., & McLaughlin, P. (2011). On the difference in swing arm kinematics between low handicap golfers and non-golfers using wireless inertial sensors. *Procedia Engineering*, *13*, 219–225. https://doi.org/10.1016/j.proeng.2011.05.076

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

LeCun, Y., Bottou, L., Orr, G. B., & Müller, K. R. (1998). Efficient backprop. In *Neural Networks: Tricks of the Trade* (Vol. 1524, pp. 9–50).

LeCun, Y., Yoshua, B., & Geoffrey, H. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, J., Tian, Q., Zhang, G., Zheng, F., Lv, C., & Wang, J. (2018). Research on hybrid information recognition algorithm and quality of golf swing. *Computers and Electrical Engineering*, 1–13. https://doi.org/10.1016/j.compeleceng.2018.02.013

Liao, W. H., Liao, Z. X., & Liu, M. J. (2003). Swimming style classification from video sequences. In Kinmen (Ed.), *16th IPPR Conference on Computer Vision, Graphics and Image Processing* (pp. 226–233). ROC.

Lu, W. L., Okuma, K., & Little, J. J. (2009). Tracking and recognizing actions of multiple hockey players using the boosted particle filter. *Image and Vision Computing*, *27*(1–2), 189–205. https://doi.org/10.1016/j.imavis.2008.02.008

Magalhaes, F. A. de, Vannozzi, G., Gatta, G., & Fantozzi, S. (2015). Wearable inertial sensors in swimming motion analysis: A systematic review. *Journal of Sports Sciences*, *33*(7), 732–745. https://doi.org/10.1080/02640414.2014.962574

Mannini, A., & Sabatini, A. M. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, *10*(2), 1154–1175. https://doi.org/10.3390/s100201154

McNamara, D. J., Gabbett, T., Blanch, P., & Kelly, L. (2017). The relationship between wearable microtechnology device variables and cricket fast bowling intensity. *International Journal of Sports Physiology and Performance*, 1–20. https://doi.org/https://doi.org/10.1123/ijspp.2016-0540

McNamara, D. J., Gabbett, T., Chapman, P., Naughton, G., & Farhart, P. (2015). The validity of microsensors to automatically detect bowling events and counts in cricket fast bowlers. *International Journal of Sports Physiology and Performance*, *10*(1), 71–75. https://doi.org/10.1123/ijspp.2014-0062

Minnen, D., Westeyn, T. L., Starner, T., Ward, J. a, & Lukowicz, P. (2006). Performance metrics and evaluation issues for continuous activity recognition. In *Proc. Int. Workshop on Performance Metrics for Intelligent Systems* (pp. 141–148). https://doi.org/10.1145/1889681.1889687

Mitchell, E., Monaghan, D., & O’Connor, N. E. (2013). Classification of sporting activities using smartphone accelerometers. *Sensors (Basel, Switzerland)*, *13*(4), 5317–5337. https://doi.org/10.3390/s130405317

Mohammed, M., Khan, M., & Bashier, E. (2016). *Machine Learning: Algorithms and Applications*. Milton: CRC Press.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Montoliu, R., Martín-Félez, R., Torres-Sospedra, O., & Martínez-Usó, A. (2015). Team activity recognition in Association football using a bag-of-words-based method. *Human Movement Science*, *41*, 165–178. https://doi.org/10.1016/j.humov.2015.03.007

Mooney, R., Corley, G., Godfrey, A., Quinlan, L. R., & ÓLaighin, G. (2015). Inertial sensor technology for elite swimming performance analysis: A systematic review. *Sensors*, *16*(1), 18. https://doi.org/10.3390/s16010018

Nibali, A., He, Z., Morgan, S., & Greenwood, D. (2017). Extraction and classification of diving clips from continuous video footage. *ArXiv, pre-print*. Retrieved from https://arxiv.org/pdf/1705.09003.pdf

O’Reilly, M. A., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017a). Classification of lunge biomechanics with multiple and individual inertial measurement units. *Sports Biomechanics*, *16*(3), 342–360. https://doi.org/10.1080/14763141.2017.1314544

O’Reilly, M. A., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017b). Technology in strength and conditioning tracking lower-limb exercises with wearable sensors. *Journal of*

Strength and Conditioning Research, *31*(6), 1726–1736.

O'Reilly, M., Caulfield, B., Ward, T., Johnston, W., & Doherty, C. (2018). Wearable Inertial Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review. *Sports Medicine*. https://doi.org/10.1007/s40279-018-0878-4

O'Reilly, M., Whelan, D., Chanialidis, C., Friel, N., Delahunt, E., Ward, T., & Caulfield, B. (2015). Evaluating squat performance with a single inertial measurement unit. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks*. IEEE. https://doi.org/10.1109/BSN.2015.7299380

O'Reilly, M., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017). Classification of deadlift biomechanics with wearable inertial measurement units. *Journal of Biomechanics, 58*, 155–161. https://doi.org/10.1080/14763141.2017.1314544

Ó Conaire, C., Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., & Buckley, J. (2010). Combining inertial and visual sensing for human action recognition in tennis. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams* (pp. 51–56). ACM. https://doi.org/10.1145/1877868.1877882

Pernek, I., Kurillo, G., Stiglic, G., & Bajcsy, R. (2015). Recognizing the intensity of strength training exercises with wearable sensors. *Journal of Biomedical Informatics*, *58*, 145–155. https://doi.org/10.1016/j.jbi.2015.09.020

Plötz, T., Hammerla, N. Y., & Olivier, P. (2011). Feature learning for activity recognition in ubiquitous computing. *International Joint Conference on Artificial Intelligence (IJCAI),* 1729.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing*, *28*(6), 976–990. https://doi.org/10.1016/j.imavis.2009.11.014

Preece, S. J., Goulermas, J. Y., Kenney, L., & Howard, D. (2009). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering*, *56*(3), 871–879. https://doi.org/10.1109/TBME.2008.2006190

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors: A review of classification techniques. *Physiological Measurement*, *30*(4), R1–R33. https://doi.org/10.1088/0967-3334/30/4/R01

Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, *42*(3), 203–231. https://doi.org/10.1023/A:1007601015854

Qaisar, S., Imtiaz, S., Glazier, P., Farooq, F., Jamal, A., Iqbal, W., & Lee, S. (2013). A method for cricket bowling action classification and analysis using a system of inertial sensors. In *International Conference on Computational Science and its Applications* (pp. 396–412). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39649-6

Ramanathan, V., Huang, J., Abu-El-Haija, S., Gorban, A., Murphy, K., & Fei-Fei, L. (2015). Detecting events and key actors in multi-person videos. https://doi.org/10.1109/CVPR.2016.332

Rassem, A., El-Beltagy, M., & Saleh, M. (2017). Cross-country skiing gears classification using deep learning. *ArXiv Preprint ArXiv:1706.08924*. Retrieved from https://arxiv.org/pdf/1706.08924v1.pdf

Ravi, D., Wong, C., Lo, B., & Yang, G.-Z. (2016). A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE Journal of Biomedical and Health Informatics*, *21*(1), 1–1. https://doi.org/10.1109/JBHI.2016.2633287

Reily, B., Zhang, H., & Hoff, W. (2017). Real-time gymnast detection and performance analysis with a portable 3D camera. *Computer Vision and Image Understanding*, *159,* 154–163. https://doi.org/10.1016/j.cviu.2016.11.006

Rindal, O. M. H., Seeberg, T. M., Tjønnås, J., Haugnes, P., & Sandbakk, Ø. (2018). Automatic classification of sub-techniques in classical cross-country skiing using a machine learning algorithm on micro-sensor data. *Sensors (Switzerland), 18*(1), 75. https://doi.org/10.3390/s18010075

Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications*, *59,* 235–244. https://doi.org/10.1016/j.eswa.2016.04.032

Saba, T., & Altameem, A. (2013). Analysis of vision based systems to detect real time goal events in soccer videos. *Applied Artificial Intelligence*, *27*(7), 656–667. https://doi.org/10.1080/08839514.2013.787779

Salman, M., Qaisar, S., & Qamar, A. M. (2017). Classification and legality analysis of bowling action in the game of cricket. *Data Mining and Knowledge Discovery, 31*(6), 1706–1734. https://doi.org/10.1007/s10618-017-0511-4

Schuldhaus, D., Zwick, C., Körger, H., Dorschky, E., Kirk, R., & Eskofier, B. M. (2015). Inertial sensor-based approach for shot/ pass classification during a soccer match. In *Proc. 21st ACM KDD Workshop on Large-Scale Sports Analytics* (pp. 1–4). Sydney, Australia.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2008). RUSBoost: Improving classification performance when training data is skewed. In *9th International Conference on Pattern Recognition* (pp. 1–4). https://doi.org/10.1109/ICPR.2008.4761297

Shah, H., Chokalingam, P., Paluri, B., & Pradeep, N. (2007). Automated stroke classification in tennis. *Image Analysis and Recognition*, 1128–1137.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: from theory to algorithms*. New York, USA: Cambridge University Press.

Sharma, M., Srivastava, R., Anand, A., Prakash, D., & Kaligounder, L. (2017). Wearable motion sensor based phasic analysis of tennis serve for performance feedback. In *2017 IEEE International Conference on Acoustics, Speech and Siginal Processing* (pp. 5945–5949). New Orleans, LA: IEEE.

Sprager, S., & Juric, M. B. (2015). *Inertial sensor-based gait recognition: A review*. *Sensors (Switzerland)* (Vol. 15). https://doi.org/10.3390/s150922089

Srivastava, R., Patwari, A., Kumar, S., Mishra, G., Kaligounder, L., & Sinha, P. (2015). Efficient characterization of tennis shots and game analysis using wearable sensors data. In *2015 IEEE Sensors- Proceedings* (pp. 1–4). Busan. https://doi.org/10.1109/ICSENS.2015.7370311

Stein, M., Janetzko, H., Lamprecht, A., Breitkreutz, T., Zimmermann, P., Goldlücke, B., … Keim, D. A. (2018). Bring it to the pitch: combining video and movement data to enhance team sport analysis. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 13–22. https://doi.org/10.1109/TVCG.2017.2745181

Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. (2017). Efficient processing of deep neural networks: A tutorial and survey. *IEEE*, *105*(2), 2295–2329. Retrieved from http://arxiv.org/abs/1703.09039

Thomas, G., Gade, R., Moeslund, T. B., Carr, P., & Hilton, A. (2017). Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, *159*, 3–18. https://doi.org/10.1016/j.cviu.2017.04.011

Titterton, D. H., & Weston, J. L. (2009). *Strapdown inertial navigation technology* (2nd ed.). Reston, VA: AIAA.

Tora, M. R., Chen, J., & Little, J. J. (2017). Classification of puck possession events in ice hockey. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 147–154). https://doi.org/10.1109/CVPRW.2017.24

Victor, B., He, Z., Morgan, S., & Miniutti, D. (2017). Continuous video to simple signals for swimming stroke detection with convolutional neural networks. *ArXiv Preprint ArXiv:1705.09894*. https://doi.org/10.1111/j.1467-8330.1974.tb00606.x

Wagner, D., Kalischewski, K., Velten, J., & Kummert, A. (2017). Activity recognition using inertial sensors and a 2-D convolutional neural network. In IEEE (Ed.), *2017 10th International Workshop on Multidimensional (nD) Systems (nDS)* (pp. 1–6). https://doi.org/10.1109/NDS.2017.8070615

Wagner, J. F. (2018). About motion measurement in sports based on gyroscopes and accelerometers - an engineering point of view. *Gyroscopy and Navigation*, *9*(1), 1–18. https://doi.org/10.1134/S2075108718010091

Ward, J. A., Lukowicz, P., & Gellersen, H.-W. (2011). Performance metrics for activity recognition. In *ACM Trans. on Intelligent Systems and Technology* (Vol. 2, pp. 111–132).

Whiteside, D., Cant, O., Connolly, M., & Reid, M. (2017). Monitoring hitting load in tennis using inertial sensors and machine learning. *International Journal of Sports Physiology and Performance*, 1–20. https://doi.org/https://doi.org/10.1123/ijspp.2016-0683

Wixted, A., Billing, D. C., & James, D. A. (2010). Validation of trunk mounted inertial sensors for analysing running biomechanics under field conditions, using synchronously collected foot contact data. *Sports Engineering*, *12*(4), 207–212. https://doi.org/10.1007/s12283-010-0043-2

Wixted, A., Portus, M., Spratford, W., & James, D. A. (2011). Detection of throwing in cricket using wearable sensors. *Sports Technology*, *4*(3–4), 134–140. https://doi.org/10.1080/19346182.2012.725409

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, *8*(7), 1341–1390. https://doi.org/10.1162/neco.1996.8.7.1391

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimisation. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82. https://doi.org/10.1023/A:1021251113462

Wundersitz, D. W., Gastin, P. B., Richter, C., Robertson, S., & Netto, K. J. (2015). Validity of a trunk-mounted accelerometer to assess peak accelerations during walking, jogging and running. *European Journal of Sport Science*, *15*(5), 382–390. https://doi.org/10.1080/17461391.2014.955131

Wundersitz, D. W., Gastin, P. B., Robertson, S., Davey, P. C., & Netto, K. J. (2015). Validation of a trunk-mounted accelerometer to measure peak impacts during team sport movements. *International Journal of Sports Medicine*, *36*(9), 742–746. https://doi.org/10.1055/s-0035-1547265

Wundersitz, D. W., Josman, C., Gupta, R., Netto, K. J., Gastin, P. B., & Robertson, S. (2015). Classification of team sport activities using a single wearable tracking device. *Journal of Biomechanics*, *48*(15), 3975–3981. https://doi.org/10.1016/j.jbiomech.2015.09.015

Yang, C. C., & Hsu, Y. L. (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, *10*(8), 7772–7788. https://doi.org/10.3390/s100807772

Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., & Shonali, K. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of the 24th International Conference on Artificial Intelligence* (pp. 3995–4001).

Yao, B., & Fei-Fei, L. (2010). Modeling mutual context of object and human pose in human-object interaction activities. In *Computer Vision and Pattern Recognition* (pp. 17–24). IEEE.

Young, C., & Reinkensmeyer, D. J. (2014). Judging complex movement performances for excellence: a principal components analysis-based technique applied to competitive diving. *Human Movement Science*, *36*, 107–122. https://doi.org/10.1016/j.humov.2014.05.009

Yu, G., Jang, Y. J., Kim, J., Kim, J. H., Kim, H. Y., Kim, K., & Panday, S. B. (2016). Potential of IMU sensors in performance analysis of professional alpine skiers. *Sensors (Switzerland)*, *16*(4), 1–21. https://doi.org/10.3390/s16040463

Zebin, T., Scully, P. J., & Ozanyan, K. B. (2016). Human Activity Recognition with Inertial Sensors Using a Deep Learning Approach. *Proc. of IEEE Sensors 2016*, (1), 1–3. https://doi.org/10.1109/ICSENS.2016.7808590

Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014). Convolutional neural networks for human activity recognition using mobile sensors. In *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services* (pp. 197–205). https://doi.org/10.4108/icst.mobicase.2014.257786

Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., & Li, Z. (2017). A review on human activity recognition using vision-based method. *Journal of Healthcare Engineering*, *2017*, 1–31. https://doi.org/10.1155/2017/3090343

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). Time series classification using multi-channels deep convolutional neural networks. In *International Conference on Web-Age Information Management* (pp. 298–310). Springer. https://doi.org/10.1007/978-3-319-08010-9_33

Zhu, G., Xu, C., Gao, W., & Huang, Q. (2006). Action recognition in broadcast tennis video. *Computer Vision in Human-Computer Interaction*, 89–98. https://doi.org/10.1007/11754336_9

Ziaeefard, M., & Bergevin, R. (2015). Semantic human activity recognition: A literature review. *Pattern Recognition*, *48*(8), 2329–2345. https://doi.org/10.1016/j.patcog.2015.03.006

Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance

Emily E. Cust[1, 2*], *Alice J. Sweeting*[1, 2], *Kevin Ball*[1] *and Sam Robertson*[1, 2]

Author details:

[1] Institute for Health and Sport (IHES), Victoria University, Melbourne, Australia

[2] Western Bulldogs Football Club, Footscray, Melbourne, Australia

Author ORCID

Sam Robertson 0000-0002-8330-0011

Alice Sweeting 0000-0002-9185-6773

Emily Cust 0000-0001-6927-6329

* Corresponding author:

Emily Cust

Email: emily.cust1@live.vu.edu.au

Institute for Health and Sport (IHES), Victoria University,

P.O. Box 14428, Melbourne, VIC 8001, Australia.

Running title:

Machine and deep learning for sport movement recognition review

.

Abstract

Objective assessment of an athlete's performance is of importance in elite sports to facilitate detailed analysis. The implementation of automated detection and recognition of sport-specific movements overcomes the limitations associated with manual performance analysis methods. The object of this study was to systematically review the literature on machine and deep learning for sport-specific movement recognition using inertial measurement unit (IMU) and, or computer vision data inputs. A search of multiple databases was undertaken. Included studies must have investigated a sport-specific movement and analysed via machine or deep learning methods for model development. A total of 52 studies met the inclusion and exclusion criteria. Data pre-processing, processing, model development and evaluation methods varied across the studies. Model development for movement recognition were predominantly undertaken using supervised classification approaches. A kernel form of the Support Vector Machine algorithm was used in 53% of IMU and 50% of vision-based studies. Twelve studies used a deep learning method as a form of Convolutional Neural Network algorithm and one study also adopted a Long Short Term Memory architecture in their model. The adaptation of experimental set-up, data pre-processing, and model development methods are best considered in relation to the characteristics of the targeted sports movement(s).

Key Words:

Sport movement classification; inertial sensors; computer vision; machine learning; performance analysis.

1. Introduction

Performance analysis in sport science has experienced considerable recent changes, due largely to access to improved technology and increased applications from computer science. Manual notational analysis or coding in sports, even when performed by trained analysts, has limitations. Such methods are typically time intensive, subjective in nature, and prone to human error and bias. Automating sport movement recognition and its application towards coding has the potential to enhance both the efficiency and accuracy of sport performance analysis. The potential automation of recognising human movements, commonly referred to as human activity recognition (HAR), can be achieved through machine or deep learning model approaches. Common data inputs are obtained from inertial measurement units (IMUs) or vision. Detection refers to the identification of a targeted instance, i.e., tennis strokes within a continuous data input signal (Bulling, Blanke, & Schiele, 2014). Recognition or classification of movements involves further interpretations and labelled predictions of the identified instance (Bulling et al., 2014; Bux, Angelov, & Habib, 2017), i.e., differentiating tennis strokes as a forehand or backhand. In machine and deep learning, a model represents the statistical operations involved in the development of an automated prediction task (LeCun, Yoshua, & Geoffrey, 2015; Shalev-Shwartz & Ben-David, 2014).

Human activities detected by inertial sensing devices and computer vision are represented as wave signal features corresponding to specific actions, which can be logged and extracted. Human movement activities are considered hierarchically structured and can be broken down to basic movements. Therefore, the context of signal use, intra-class variability, and inter-class similarity between activities require consideration during experimental set-up and model development. Wearable IMUs contain a combination of accelerometer, gyroscope, and magnetometer sensors measuring along one to three axes. These sensors quantify acceleration, angular velocity, and the direction and orientation of travel respectively (Gastin, McLean, Breed, & Spittle, 2014). These sensors can capture repeated movement patterns during sport training and competitions (Camomilla, Bergamini, Fantozzi, & Vannozzi, 2018; Chambers, Gabbett, Cole, & Beard, 2015; J. F. Wagner, 2018). Advantages include being wireless, lightweight and self-contained in operation. Inertial measurement units have been utilised in quantifying physical output and tackling impacts in Australian Rules football (Gastin et al., 2014; Gastin, McLean, Spittle, & Breed, 2013) and rugby

(Gabbett, Jenkins, & Abernethy, 2012, 2011; Howe, Aughey, Hopkins, Stewart, & Cavanagh, 2017; Hulin, Gabbett, Johnston, & Jenkins, 2017). Other applications include swimming analysis (Mooney, Corley, Godfrey, Quinlan, & ÓLaighin, 2015), golf swing kinematics (Lai, Hetchl, Wei, Ball, & McLaughlin, 2011), over-ground running speeds (Wixted, Billing, & James, 2010), full motions in alpine skiing (Yu et al., 2016); and the detection and evaluation of cricket bowling (McNamara, Gabbett, Blanch, & Kelly, 2017; McNamara, Gabbett, Chapman, Naughton, & Farhart, 2015; Wixted, Portus, Spratford, & James, 2011).

Computer vision has applications for performance analysis including player tracking, semantic analysis, and movement analysis (Stein et al., 2018; Thomas, Gade, Moeslund, Carr, & Hilton, 2017). Automated movement recognition approaches require several pre-processing steps including athlete detection and tracking, temporal cropping and targeted action recognition, which are dependent upon the sport and footage type (Barris & Button, 2008; Saba & Altameem, 2013; Thomas et al., 2017). Several challenges including occlusion, viewpoint variations, and environmental conditions may impact results, depending on the camera set-up (Poppe, 2010; Zhang et al., 2017). Developing models to automate sports-vision coding may improve resource efficiency and reduce feedback times. For example, coaches and athletes involved in time-intensive notational tasks, including post-swim race analysis, may benefit from rapid objective feedback before the next race in the event program (Liao, Liao, & Liu, 2003; Victor, He, Morgan, & Miniutti, 2017). For detecting and recognising movements, body worn sensor signals do not suffer from the same environmental constraints and stationary set-up of video cameras. Furthermore, multiple sensors located on different body segments have been argued to provide more specific signal representations of targeted movements (J. B. Yang, Nguyen, San, Li, & Shonali, 2015). But it is not clear if this is solely conclusive, and the use of body worn sensors in some sport competitions may be impractical or not possible.

Machine learning algorithms learn from data input for automated model building and perform tasks without being explicitly programmed. The algorithm goal is to output a response function $h\sigma(\bar{x})$ that will predict a ground truth variable y from an input vector of variables $\bar{x}$. Models are run for classification techniques to predict a target class (Kotsiantis, Zaharakis, & Pintelas, 2007), or regression to predict discrete or continuous values. Models are aimed at finding an optimal set of parameters σ to describe the response function, and then make predictions on unseen unlabelled data input. Within these, model training approaches can generally run as supervised learning, unsupervised learning or semi-supervised learning (Mohammed, Khan, & Bashier, 2016; Sze, Chen, Yang, & Emer, 2017).

Processing raw data is limited for conventional machine learning algorithms, as they are unable to effectively be trained on abstract and high-dimensional data that is inconsistent, contains missing values or noisy artefacts (Bux et al., 2017; Kautz, 2017). Consequently, several pre-processing stages are required to create a suitable data form for input into the classifier algorithm (Figo, Diniz, Ferreira, & Cardoso, 2010). Filtering (Figo et al., 2010; Wundersitz, Gastin, Robertson, Davey, & Netto, 2015), window capture durations (Mitchell, Monaghan, & O'Connor, 2013; Preece, Goulermas, Kenney, & Howard, 2009; Wundersitz, Josman, et al., 2015), and signal frequency cut-offs (Wundersitz, Gastin, Richter, Robertson, & Netto, 2015; Wundersitz, Gastin, Robertson, et al., 2015) are common techniques applied prior to data prior to dynamic human movement recognition. Well-established filters for processing motion signal data include the Kalman filter (Kautz, 2017; Titterton & Weston, 2009; D. Wagner, Kalischewski, Velten, & Kummert, 2017) and a Fourier transform filter (Preece, Goulermas, Kenney, Howard, et al., 2009) such as a fast Fourier transform (Kapela, Świetlicka, Rybarczyk, Kolanowski, & O'Connor, 2015; Preece, Goulermas, Kenney, & Howard, 2009). Near real-time processing benefits from reducing memory requirements, computational demands, and essential bandwidth during whole model implementation. Signal feature extraction and selection favours faster processing by reducing the signals to the critical features that can discriminate the targeted activities (Bulling et al., 2014). Feature extraction involves identifying the key features that help maximise classifier success, and removing features that have minimal impact in the model (Mannini & Sabatini, 2010). Thus, feature selection involves constructing data representations in subspaces with reduced dimensions. These identified variables are represented in a compact feature variable (Mannini & Sabatini, 2010). Common methods include principal component analysis (PCA) (Gløersen, Myklebust, Hallén, & Federolf, 2018; Young & Reinkensmeyer, 2014), vector coding techniques (Hafer & Boyer, 2017) and empirical cumulative distribution functions (ECDF) (Plötz, Hammerla, & Olivier, 2011). An ECDF approach has been shown to be advantageous over PCA as it derives representations of raw input independent of the absolute data ranges, whereas PCA is known to have reduced performance when the input data is not properly normalised (Plötz et al., 2011). For further detailed information on the acquisition, filtering and analysis of IMU data for sports application and vision-based human activity recognition, see (Kautz, 2017) and (Bux et al., 2017), respectively.

Deep learning is a division of machine learning, characterised by deeper neural network model architectures and are inspired by the biological neural networks of the human brain (Bengio, 2013; LeCun et al., 2015; Sze et al., 2017). The deeper hierarchical models create a profound architecture of multiple hidden layers based on representative learning with several processing and abstraction layers (Bux et al., 2017; J. B. Yang et al., 2015). These computational models allow data input features to be automatically extracted from raw data and transformed to handle unstructured data, including vision (LeCun et al., 2015; Ravi, Wong, Lo, & Yang, 2016). This direct input avoids several processing steps required in machine learning during training and testing, therefore reducing overall computational times. A current key element within deep learning is backpropagation (Hecht-Nielsen, 1989; LeCun, Bottou, Orr, & Müller, 1998). Backpropagation is a fast and computationally efficient algorithm, using gradient descent, that allows training deep neural networks to be tractable (Sze et al., 2017). Human activity recognition has mainly been performed using conventional machine learning classifiers. Recently, deep learning techniques have enhanced the bench mark and applications for IMUs (Kautz et al., 2017; Ravi et al., 2016; Ronao & Cho, 2016; J. B. Yang et al., 2015; Zebin, Scully, & Ozanyan, 2016; Zeng et al., 2014) and vision (Ji, Yang, Yu, & Xu, 2013; Karpathy et al., 2014a; Krizhevsky, Sutskever, & Hinton, 2012; Nibali, He, Morgan, & Greenwood, 2017) in human movement recognition producing more superior model performance accuracy.

The objective of this study was to systematically review the literature investigating sport-specific automated movement detection and recognition. The review focusses on the various technologies, analysis techniques and performance outcome measures utilised. There are several reviews within this field that are sensor-based including wearable IMUs for lower limb biomechanics and exercises (Fong & Chan, 2010; M. O'Reilly, Caulfield, Ward, Johnston, & Doherty, 2018), swimming analysis (Magalhaes, Vannozzi, Gatta, & Fantozzi, 2015; Mooney et al., 2015), quantifying sporting movements (Chambers et al., 2015) and physical activity monitoring (C. C. Yang & Hsu, 2010). A recent systematic review has provided an evaluation on the in-field use of inertial-based sensors for various performance evaluation applications (Camomilla et al., 2018). Vision-based methods for human activity recognition (Aggarwal & Xia, 2014; Bux et al., 2017; Ke et al., 2013; Zhang et al., 2017), semantic human activity recognition (Ziaeefard & Bergevin, 2015) and motion analysis in sport (Barris & Button, 2008) have also been reviewed. However, to date, there is no systematic review across sport-specific movement detection and recognition via machine or deep learning. Specifically, incorporating IMUs and vision-based data input, focussing on in-field applications as opposed to laboratory-based protocols and detailing the analysis and machine learning methods used.

Considering the growth in research and potential field applications, such a review is required to understand the research area. This review aims to characterise the evolving techniques and inform researchers of possible improvements in sports analysis applications. Specifically: 1) What is the current scope for IMUs and computer vision in sport movement detection and recognition? 2) Which methodologies, inclusive of signal processing and model learning techniques, have been used to achieve sport movement recognition? 3) Which evaluation methods have been used in assessing the performance of these developed models?

2. Methods

2.1 Search strategy

The preferred PRISMA recommendations (Moher, Liberati, Tetzlaff, Altman, & Group, 2009) for systematic reviews were used. A literature search was undertaken by the first author on the following databases; IEEE Xplore, PubMed, ScienceDirect, Scopus, Academic Search Premier, and Computer and Applied Science Complete. The searched terms were categorised in order to define the specific participants, methodology and evaluated outcome measure in-line with the review aims. Searches used a combination of key words with AND/OR phrases which are detailed in Table 1. Searches were filtered for studies from January 2000 to May 2018 as no relevant studies were identified prior to this. Further studies were manually identified from the bibliographies of database-searched studies identified from the abstract screen phase, known as snowballing. Table 2 provides the inclusion and exclusion criteria of this review.

*****Table 1 near here: Key word search term string*s* per database *****

*****Table 2 near here: Inclusion and exclusion criteria*****

2.2 Data extraction

The first author extracted and collated the relevant information from the full manuscripts identified for final review. A total of 18 parameters were extracted from the 52 research studies, including the title, author, year of publication, sport, participant details, sport movement target(s), device specifications, device sample frequency, pre-processing methods, processing methods, feature selected, feature extraction, machine learning model used, model evaluation, model performance accuracy, validation method, samples collected, and computational information. A customised Microsoft Excel^{TM} spreadsheet was developed to categorise the relevant extracted information from each study. Participant characteristics of number of participants, gender, and competition level, then if applicable a further descriptor specific to a sport, for example, ‘medium-paced cricket bowler’. Athlete and participant experience level was categorised as written in the corresponding study to avoid misrepresentations. The age of participants was not considered an important characteristic required for model development. The individual ability in which the movement is performed accounts for the discriminative signal features associated with the movements. For the purposes of this review, a sport-specific movement was defined from a team or individual sport, and training activities associated with a particular sport. For example, weight-lifting as strength training, recognised under the Global Association of International Sports Federations. The targeted sports and specific movements were defined for either detection or recognition. Model development techniques used included pre-processing methods to transform data to a more suitable form for analysis, processing stages to segment data for identified target activities, feature extraction and selections techniques, and the learning algorithm(s). Model evaluation measures extracted were the model performance assessment techniques used, ground-truth validation comparison, number of data samples collected, and the model performance outcomes results reported. If studies ran multiple experiments using several algorithms, only the superior algorithm and relevant results were reported as the best method. This was done so in the interest of concise reporting to highlight favourable method approaches (Sprager & Juric, 2015). Any further relevant results or information identified from the studies was included as a special remark (Sprager & Juric, 2015). Hardware and specification information extracted included the IMU or video equipment used, number of units, attachment of sensors (IMUs), sample frequency, and sensor data types used in analysis (IMUs). Studies identified and full data extracted were reviewed by a second author.

3. Results

An outline of the search results and study exclusions has been provided in Fig 1. Of the initial database search which identified 4885 results, a final 52 studies met criteria for inclusion in this review. Of these, 29 used IMUs and 22 were vision-based. One study (Ó Conaire et al., 2010) used both sensors and vision for model development separately then together via data fusion. Tables 3 - 8 provide a description of the characteristics of the reviewed studies, detailed in the following sections.

***** Fig 1 near here: PRISMA flow diagram *****

3.1 Experimental design

A variety of sports and their associated sport-specific movements were investigated, implementing various experimental designs as presented in Tables 5 and 7. Across the studies, sports reported were tennis (n = 10), cricket (n = 3), weightlifting or strength training (n = 6), swimming (n = 4), skateboarding (n = 2), ski jumping (n = 2), snowboarding (n = 1), golf (n = 4), volleyball (n = 2), rugby (n = 2), ice hockey (n = 2), gymnastics (n = 2), karate (n = 1), basketball (n = 3), Gaelic football (n = 1), hurling (n = 1), boxing (n = 2), running (n = 2), diving (n = 1), squash (n = 1), badminton (n = 1), cross-country skiing (n = 2) and soccer (n = 4). The Sports 1-M dataset (Karpathy et al., 2014b) was also reported, which consists of 1,133,158 video URLs annotated automatically with 487 sport labels using the YouTube Topic API. A dominant approach was the classification of main characterising actions for each sport. For example, serve, forehand, backhand strokes in tennis (Connaghan et al., 2011; Kos & Kramberger, 2017; Ó Conaire et al., 2010; Shah, Chokalingam, Paluri, & Pradeep, 2007; Srivastava et al., 2015), and the four competition strokes in swimming (Jensen, Blank, Kugler, & Eskofier, 2016; Jensen, Prade, & Eskofier, 2013; Liao et al., 2003; Victor et al., 2017). Several studies further classified sub-categories of actions. For example, three further classes of the two main classified snowboarding trick types Grinds and Airs (Groh, Fleckenstein, & Eskofier, 2016), and further classifying the main tennis stroke types as either flat, topspin or slice

(Srivastava et al., 2015). Semantic descriptors were reported for classification models that predicted athlete training background, experience and fatigue level. These included running (Buckley et al., 2017; Kobsar, Osis, Hettinga, & Ferber, 2014), rating of gymnastic routines (Reily, Zhang, & Hoff, 2017), soccer pass classification based on its quality (Horton, Gudmundsson, Chawla, & Estephan, 2014), cricket bowling legality (Qaisar et al., 2013; Salman, Qaisar, & Qamar, 2017), ski jump error analysis (Brock & Ohgi, 2017; Brock, Ohgi, & Lee, 2017) and strength training technique deviations (M. A. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017a; M. O'Reilly et al., 2015; M. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017). One approach (Yao & Fei-Fei, 2010), encoded the mutual context of human pose and sporting equipment using semantics, to facilitate the detection and classification of movements including a cricket bat and batsman coupled movements.

Total participant numbers for IMU-based studies ranged from one (Qaisar et al., 2013) to 30 (Kautz et al., 2017). Reported data individual instance sample sizes for sensor studies ranged from 150 (Salman et al., 2017) to 416, 737 (Rassem, El-Beltagy, & Saleh, 2017). Vision-based studies that explicitly reported total participant details ranged from five (Ó Conaire et al., 2010) to 40 (Victor et al., 2017). Vision dataset sample sizes varied across studies, from 50 individual action clips (Liao et al., 2003) to 15, 000 (Victor et al., 2017). One study (Karpathy et al., 2014a) used the publicly available Sports-1M, as previously described. Vision-based studies also reported datasets in total time, 10.3 hours (Bertasius, Park, Yu, & Shi, 2017), 3 hours (Montoliu, Martín-Félez, Torres-Sospedra, & Martínez-Usó, 2015), 1, 500 minutes (Shah et al., 2007), and 50 hours (Kapela et al., 2015), and by frame numbers, 6, 035 frames (Zhu, Xu, Gao, & Huang, 2006) and 10, 115 frames (Reily et al., 2017).

3.2 Inertial measurement unit specifications

A range of commercially available and custom-built IMUs were used in the IMU-based studies (n= 30), as presented in Table 3. Of these, 23% reported using a custom-built sensor. Of the IMU-based studies, the number of sensors mounted or attached to each participant or sporting equipment piece ranged from one to nine. The majority of studies (n= 22) provided adequate details of sensor specifications including sensor type, axes, measurement range, and sample rate used. At least one characteristic of sensor measurement range or sample rate used in data collection was missing from eight studies. All studies used triaxial sensors and collected accelerometer data. For analysis and model development, individual sensor data consisted of only accelerometer data (n = 8), both accelerometer and gyroscope data (n = 15), and accelerometer, gyroscope and magnetometer data (n = 7). The individual sensor measurement ranges reported for accelerometer were ± 1.5 g to ± 16 g, gyroscope ± 500 ◦/s to ± 2000 ◦/s, magnetometer ± 1200 µT or 1.2 to 4 Ga. Individual sensor sample rates ranged from 10 Hz to 1000 Hz for accelerometers, 10 Hz to 500 Hz for gyroscopes and 50 Hz to 500 Hz for magnetometers.

***** Table 3 near here*****

3.3 Vision capture specification

Several experimental set-ups and specifications were reported in the total 23 vision-based studies (Table 4). Modality was predominately red, green, blue (RGB) cameras. Depth cameras were utilised (Kasiri-Bidhendi, Fookes, Morgan, Martin, & Sridharan, 2015; Kasiri, Fookes, Sridharan, & Morgan, 2017; Reily et al., 2017), which add depth perception for 3-dimensional image mapping. Seven studies clearly reported the use of a single camera set-up (Couceiro, Dias, Mendes, & Araújo, 2013; Díaz-Pereira, Gómez-Conde, Escalona, & Olivieri, 2014; Hachaj, Ogiela, & Koptyra, 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Nibali et al., 2017; Reily et al., 2017). One study reported 16 stationary positioned cameras at a ‘bird’s eye view’ (Montoliu et al., 2015), and Ó Conaire et al. (2010) reported the use of one overhead and 8 stationary cameras around a tennis court baseline, although data from two cameras were only used in final analysis due to occlusion issues. Sample frequency and, or pixel resolution were reported in seven of the studies (Couceiro et al., 2013; Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Montoliu et al., 2015; Victor et al., 2017; Zhu et al., 2006), with sample frequencies ranging from 30 Hz to 210 Hz.

***** Table 4 near here*****

3.4 Inertial measurement unit recognition model development methods

Key stages of model development from data pre-processing to recognition techniques for IMU-based studies are presented in Table 5. Data pre-processing filters were reported as either a low-pass filter (n = 7) (Adelsberger & Tröster, 2013; Buckley et al., 2017; Kelly, Coughlan, Green, & Caulfield, 2012; M. A. O’Reilly et al., 2017a; M. O’Reilly et al., 2015, 2017; Rindal, Seeberg, Tjønnås, Haugnes, & Sandbakk, 2018), high-pass filter (n = 2) (Kautz et al., 2017; Schuldhaus et al., 2015), or calibration with a filter (Salman et al., 2017). Processing methods were reported in 67% of the IMU-based studies (Adelsberger & Tröster, 2013; Anand, Sharma, Srivastava, Kaligounder, & Prakash, 2017; Brock et al., 2017; Buckley et al., 2017; Buthe, Blanke, Capkevics, & Tröster, 2016; Groh et al., 2016; Groh, Fleckenstein, Kautz, & Eskofier, 2017; Groh, Kautz, & Schuldhaus, 2015; Jensen et al., 2016, 2015; Jiao, Wu, Bie, Umek, & Kos, 2018; Kautz et al., 2017; Kobsar et al., 2014; M. A. O’Reilly et al., 2017a; M. O’Reilly et al., 2017; Ó Conaire et al., 2010; Pernek, Kurillo, Stiglic, & Bajcsy, 2015; Qaisar et al., 2013; Salman et al., 2017; Schuldhaus et al., 2015). Methods included, calibration of data (Groh et al., 2016, 2017; Jensen et al., 2015; Qaisar et al., 2013), a one-second window centred around identified activity peaks in the signal (Adelsberger & Tröster, 2013; Schuldhaus et al., 2015), temporal alignment (Pernek et al., 2015), normalisation (Ó Conaire et al., 2010), outlier adjustment (Kobsar et al., 2014) or removal (Salman et al., 2017), and sliding windows ranging from one to 3.5 seconds across the data (Jensen et al., 2016). The three studies that investigated trick classification in skateboarding (Groh et al., 2017, 2015) and snowboarding (Groh et al., 2016) corrected data for different rider board stance styles, termed Regular or Goofy, by inverting signal axes.

Movement detection methods were specifically reported in 16 studies (Adelsberger & Tröster, 2013; Anand et al., 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015, Jensen et al., 2013, 2015; Kautz et al., 2017; Kelly et al., 2012; Kos & Kramberger, 2017; Ó Conaire et al., 2010; Rindal et al., 2018; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside, Cant, Connolly, & Reid, 2017). Detection methods included thresholding (n = 5), windowing segmenting (n = 4), and a combination of threshold and windowing techniques (n = 5).

Signal feature extraction techniques were reported in 80% of the studies, with the number of feature parameters in a vector ranging from a vector of normalised X, Y, Z accelerometer signals (Ó Conaire et al., 2010) to 240 features (M. A. O’Reilly et al., 2017a). Further feature selection to reduce the dimensionality of the feature vector was used in 11 studies. Both feature extraction and selection methods varied considerably across the literature (Table 5).

Algorithms trialled for movement recognition were diverse across the literature (Table 5). Supervised classification using a kernel form of Support Vector Machine (SVM) was most prevalent (n = 16) (Adelsberger & Tröster, 2013; Brock & Ohgi, 2017; Brock et al., 2017; Buckley et al., 2017; Buthe et al., 2016; Groh et al., 2016, 2017, 2015; Jensen et al., 2016; Kautz et al., 2017; Kelly et al., 2012; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside et al., 2017). The next highest tested were Naïve Bayesian (NB) (n = 8) (Buckley et al., 2017; Connaghan et al., 2011; Groh et al., 2016, 2017, 2015; Kautz et al., 2017; Salman et al., 2017; Schuldhaus et al., 2015) and k-Nearest Neighbour (kNN) (n = 8) (Buckley et al., 2017; Groh et al., 2016, 2017, 2015; Kautz et al., 2017; Ó Conaire et al., 2010; Salman et al., 2017; Whiteside et al., 2017), followed by Random Forests (RF) (n = 7) (Buckley et al., 2017; Groh et al., 2017; Kautz et al., 2017; M. A. O'Reilly et al., 2017a; M. O'Reilly et al., 2017; Salman et al., 2017; Whiteside et al., 2017). Supervised learning algorithms were the most common (n = 29). One study used an unsupervised discriminative analysis approach for detection and classification of tennis strokes (Kos & Kramberger, 2017). Five IMU-based study investigated a deep learning approach including using Convolutional Neural Networks (CNN) (Anand et al., 2017; Brock et al., 2017; Jiao et al., 2018; Kautz et al., 2017; Rassem et al., 2017) and Long Short Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) architectures (Rassem et al., 2017; Sharma, Srivastava, Anand, Prakash, & Kaligounder, 2017). In order to assess the effectiveness of the various classifiers from each study, model performance measures quantify and visualise the predictive performance as reported in the following section.

***** Table 5 near here*****

3.5 Inertial measurement unit recognition model evaluation

Reported performance evaluations of developed models across the IMU-based studies are shown in Table 6. Classification accuracy, as a percentage score for the number of correct predictions by total number of predictions made, was the main model evaluation measure (n = 24). Classification accuracies across studies ranged between 52% (Brock & Ohgi, 2017) to 100% (Buckley et al., 2017). Generally, the reported highest accuracy for a specific movement was ≥ 90% (n = 17) (Adelsberger & Tröster, 2013; Anand et al., 2017; Buckley et al., 2017; Connaghan et al., 2011; Groh et al., 2015; Jensen et al., 2013; Jiao et al., 2018; Kobsar et al., 2014; Kos & Kramberger, 2017; M. A. O'Reilly et al., 2017a; Ó Conaire et al., 2010; Pernek et al., 2015; Qaisar et al., 2013; Rindal et al., 2018;

Schuldhaus et al., 2015; Srivastava et al., 2015; Whiteside et al., 2017) and ≥ 80% to 90% (n = 7) (Brock & Ohgi, 2017; Brock et al., 2017; Groh et al., 2017; Jensen et al., 2016; M. O'Reilly et al., 2015, 2017; Salman et al., 2017). As an estimate of the generalised performance of a trained model on $n - x$ samples, a form of leave-one-out cross validation (LOO-CV) was used in 47% of studies (Buthe et al., 2016; Groh et al., 2016, 2017, 2015, Jensen et al., 2016, 2013; Kobsar et al., 2014; M. O'Reilly et al., 2015, 2017; Ó Conaire et al., 2010; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015). Precision, specificity and sensitivity (also referred to as recall) evaluations were derived for detection (n = 6) and classification models (n = 10). Visualisation of prediction results in the form of a confusion matrix featured in six studies (Buthe et al., 2016; Groh et al., 2017; Kautz et al., 2017; Pernek et al., 2015; Rindal et al., 2018; Whiteside et al., 2017).

***** Table 6 near here*****

3.6 Vision recognition model development methods

Numerous processing and recognition methods featured across the vision-based studies to transform and isolated relevant input data (Table 7). Pre-processing stages were reported in 14 of studies, and another varied 13 studies also provided details of processing techniques. Signal feature extraction and feature selection methods used were reported in 78% of studies.

Both machine (n = 16) and deep learning (n = 7) algorithms were used to recognise movements from vision data. Of these, a kernel form of the SVM algorithm was most common in the studies (n = 10) (Couceiro et al., 2013; Horton et al., 2014; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Li et al., 2018; Montoliu et al., 2015; M. A. O'Reilly, Whelan, Ward, Delahunt, & Caulfield, 2017b; Ó Conaire et al., 2010; Reily et al., 2017; Shah et al., 2007; Zhu et al., 2006). Other algorithms included kNN (n = 3) (Díaz-Pereira et al., 2014; Montoliu et al., 2015; Ó Conaire et al., 2010), decision tree (DT) (n = 2) (Kapela et al., 2015; Liao et al., 2003), RF (n = 2) (Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017), and Multilayer Perceptron (MLP) (n = 2) (Kapela et al., 2015; Montoliu et al., 2015). Deep learning was investigated in seven studies (Bertasius et al., 2017; Ibrahim, Muralidharan, Deng, Vahdat, & Mori, 2016; Karpathy et al., 2014a; Nibali et al., 2017;

Ramanathan et al., 2015; Tora, Chen, & Little, 2017; Victor et al., 2017) of which used CNNs or LSTM RNNs as the core model structure.

***** Table 7 near here*****

3.7 Vision recognition model evaluation

Performance evaluation methods and results for vision-based studies are reported in Table 8. As with IMU-based studies, classification accuracy was the common method for model evaluations, featured in 61%. Classification accuracies were reported between 60.9% (Karpathy et al., 2014a) and 100% (Hachaj et al., 2015; Nibali et al., 2017). In grouping the reported highest accuracies for a specific movement that were ≥ 90% (n = 9) (Hachaj et al., 2015; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Li et al., 2018; Montoliu et al., 2015; Nibali et al., 2017; Ó Conaire et al., 2010; Reily et al., 2017; Shah et al., 2007), and ≥ 80% to 90% (n = 2) (Horton et al., 2014; Yao & Fei-Fei, 2010). A confusion matrix as a visualisation of model prediction results was used in nine studies (Couceiro et al., 2013; Hachaj et al., 2015; Ibrahim et al., 2016; Karpathy et al., 2014a; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Lu, Okuma, & Little, 2009; Shah et al., 2007; Tora et al., 2017). Two studies assessed and reported their model computational average speed (Lu et al., 2009) and time (Reily et al., 2017).

***** Table 8 near here*****

4 Discussion

The aim of this systematic review was to evaluate the use of machine and deep learning for sport-specific movement recognition from IMUs and, or computer vision data inputs. Overall, the search yielded 52 studies, categorised as 29 which used IMUs, 22 vision-based and one study using both IMUs and vision. Automation or semi-automated sport movement recognition models working in near-real time is of particular interest to avoid the error, cost and time associated with manual methods. Evident in the literature, models are trending towards the potential to provide optimised objective assessments of athletic movement for technical and tactical evaluations. The majority of studies achieved favourable movement recognition results for the main characterising actions of a sport, with several studies exploring further applications such as an automated skill quality evaluation or judgement scoring, for example automated ski jump error evaluation (Brock et al., 2017).

Experimental set-up of IMU placement and numbers assigned per participant varied between sporting actions. The sensor attachment locations set by researchers appeared dependent upon the specific sporting conditions and movements, presumably to gain optimal signal data. Proper fixation and alignment of the sensor axes with limb anatomical axes is important in reducing signal error (Fong & Chan, 2010). The attachment site hence requires a biomechanical basis for accuracy of the movement being targeted to obtain reliable data. Single or multiple sensor use per person also impacts model development trade-off between accuracy, analysis complexity, and computational speed or demands. In tennis studies, specificity whilst using a single sensor was demonstrated by mounting the IMU on the wrist or forearm of the racquet arm (Connaghan et al., 2011; Kos & Kramberger, 2017; Srivastava et al., 2015; Whiteside et al., 2017). A single sensor may also be mounted in a low-profile manner on sporting equipment (Groh et al., 2016, 2017, 2015; Jensen et al., 2015). Unobtrusive use of a single IMU to capture generalised movements across the whole body was demonstrated, with an IMU mounted on the posterior head in swimming (Jensen et al., 2016, 2013), lower back during running (Kobsar et al., 2014), and between the shoulder blades in rugby union (Kelly et al., 2012).

The majority of vision-based studies opted for a single camera set-up of RGB modality. Data output from a single camera as opposed to multiple minimises the volume of data to process, therefore reducing computational effort. However, detailed features may go uncaptured, particularly in team sport competition which consists of multiple individuals participating in the capture space at one time. In contrast, a multiple camera set-up reduces limitations including occlusion and viewpoint variations. However, this may also increase the complexity of the processing and model computational stages. Therefore, a trade-off between computational demands and movement recording accuracy often needs to be made. As stated earlier, the placement of cameras needs to suit the biomechanical nature of the targeted movement and the environment situated in. Common camera capture systems used in sports science research such as Vicon Nexus (Oxford, UK) and OptiTrack (Oregon, USA) were not present in this review. As this review targeted studies investigating during on-field or in-situation sporting contexts, efficiency in data collection is key for routine applications in training and competition. A simple portable RGB camera is easy to set-up in a dynamic and changing environment, such as different soccer pitches, rather than a multiple capture system such as Vicon that requires calibrated precision and are substantially more expensive.

Data acquisition and type from an IMU during analysis appears to influence model trade-off between accuracy and computational effort of performance. The use of accelerometer, gyroscope or magnetometer data may depend upon the movement properties analysed. Within tennis studies, gyroscope signals were the most efficient at discriminating between stroke types (Buthe et al., 2016; Kos & Kramberger, 2017) and detecting an athlete's fast feet court actions (Buthe et al., 2016). In contrast, accelerometer signals produced higher classification accuracies in classifying tennis stroke skills levels (Connaghan et al., 2011). The authors expected lower gyroscope classification accuracies as temporal orientation measures between skill levels of tennis strokes will differ (Connaghan et al., 2011). Conversely, data fusion from all three individual sensors resulted in a more superior model for classifying advanced, intermediate and novices tennis player strokes (Connaghan et al., 2011). Fusion of accelerometer and vision data also resulted in a higher classification accuracy for tennis stroke recognition (Ó Conaire et al., 2010).

Supervised learning approaches were dominant across IMU and vision-based studies. This is a method which involves a labelled ground truth training dataset typically manually annotated by sport analysts. Labelled data instances were recorded as up to 15, 000 for vision-based (Victor et al., 2017) and 416, 737 for sensor-based (Rassem et al., 2017) studies. Generation of a training data set for supervised learning can be a tedious and labour-intensive task. It is further complicated if multiple sensors or cameras are incorporated for several targeted movements. A semi-supervised or unsupervised learning approach may be advantageous as data labelling is minimal or not required, potentially reducing human errors in annotation. An unsupervised approach could suit specific problems to explain key data features, via clustering (Mohammed et al., 2016; Sze et al., 2017). Results computed by an unsupervised model (Kos, Ženko, Vlaj, & Kramberger, 2016) for tennis serve, forehand and backhand stroke classification compared favourbaly well against a proposed supervised approach (Connaghan et al., 2011).

Recognition of sport-specific movements was primarily achieved using conventional machine learning approaches, however nine studies implemented deep learning algorithms. It is expected that future model developments will progressively feature deep learning approaches due to development of better hardware, and the advantages of more efficient model learning on large data inputs (Sze et al., 2017). Convolutional Neural networks (CNN) (LeCun, Bottou, Bengio, & Haffner, 1998) were the core structure of five of the seven deep learning study models. Briefly, convolution applies several filters, known as kernels, to automatically extract features from raw data inputs. This process works under four key ideas to achieve optimised results: local connection, shared weights, pooling and applying several layers (LeCun et al., 2015; J. B. Yang et al., 2015). Machine learning classifiers modelled with generic hand-crafted features, were compared against a CNN for classifying nine beach volleyball actions using IMUs (Kautz et al., 2017). Unsatisfactory results were obtained from the machine learning model, and the CNN markedly achieved higher classification accuracies (Kautz et al., 2017). The CNN model produced the shortest overall computation times, requiring less computational effort on the same hardware (Kautz et al., 2017). Vision-based CNN models have also shown favourable results when compared to a machine learning study baseline (Karpathy et al., 2014a; Nibali et al., 2017; Victor et al., 2017). Specifically, consistency between a swim stroke detection model for continuous videos in swimming which was then applied to tennis strokes with no domain-specific settings introduced (Victor et al., 2017). The authors of this training approach (Victor et al., 2017) anticipate that this could be applied to train separate models for other sports movement detection as the CNN model demonstrated the ability to learn to process continuous videos into a 1-D signal with the signal peaks corresponding to arbitrary events. General human activity recognition using CNN have shown to be a superior approach over conventional machine learning algorithms using both IMUs (Ravi et al., 2016; J. B. Yang et al., 2015; Zebin et al., 2016; Zeng et al., 2014; Zheng, Liu, Chen, Ge, & Zhao, 2014) and computer vision (Ji et al., 2013; Krizhevsky et al., 2012; LeCun et al., 2015). As machine learning algorithms extract heuristic features requiring domain knowledge, this creates shallower features which can make it harder to infer high-level and context aware activities (J. B. Yang et al., 2015). Given the previously described advantages of deep learning algorithms which apply to CNN, and the recent results of deep learning, future model developments may benefit from exploring these methods in comparison to current bench mark models.

Model performance outcome metrics quantify and visualise the error rate between the predicted outcome and true measure. Comparatively, a kernel form of an SVM was the most common classifier implemented and produced the strongest machine learning approach model prediction accuracies across both IMU (Adelsberger & Tröster, 2013; Brock & Ohgi, 2017; Buthe et al., 2016; Groh et al., 2016, 2017, 2015; Jensen et al., 2016; Pernek et al., 2015; Salman et al., 2017; Schuldhaus et al., 2015; Whiteside et al., 2017) and vision-based study designs (Horton et al., 2014; Kasiri-Bidhendi et al., 2015; Kasiri et al., 2017; Li et al., 2018; Reily et al., 2017; Shah et al., 2007; Zhu et al., 2006). Classification accuracy was the most common reported measure followed by confusion matrices, as ways to clearly present prediction results and derive further measures of performance. Further measures included sensitivity (also called recall), specificity and precision, whereby results closer to 1.0 indicate superior model performance, compared to 0.0 or poor model performance. The F1-score (also called a F-measure or F-score) conveys the balances between the precision and sensitivity of a model. An in-depth analysis performance metrics specific to human activity recognition is located elsewhere (Minnen, Westeyn, Starner, Ward, & Lukowicz, 2006; Ward, Lukowicz, & Gellersen, 2011). Use of specific evaluation methods depends upon the data type. Conventional performance measures of error rate are generally unsuitable for models developed from skewed training data (Provost & Fawcett, 2001). Using conventional performance measures in this context will only take the default decision threshold on a model trained, if there is an uneven class distribution this may lead to imprecision (Provost & Fawcett, 2001; Seiffert, Khoshgoftaar, Van Hulse, & Napolitano, 2008). Alternative evaluators including Receiver Operating Characteristics (ROC) curves and its single numeric measure, Area Under ROC Curve (AUC), report model performances across all decision thresholds (Seiffert et al., 2008). Making evaluations between study methodology have inherent complications due to each formulating their own experimental parameter settings, feature vectors and training algorithms for movement recognition. The No-Free-Lunch theorems are important deductions in the formation of models for supervised machine learning (David H. Wolpert, 1996), and search and optimisation algorithms (D H Wolpert & Macready, 1997). The theorems broadly reference that there is no ‘one model’ that will perform optimally across all recognition problems. Therefore, experiments with multiple model development methods for a particular problem is recommended. The use of prior knowledge about the task should be implemented to adapt the model input and model parameters in order to improve overall model success (Shalev-Shwartz & Ben-David, 2014).

Acquisition of athlete specific information, including statistics on number, type and intensity of actions, may be of use in the monitoring of athlete load. Other potential applications include personalised movement technique analysis (M. O'Reilly et al., 2017), automated performance evaluation scoring (Reily et al., 2017) and team ball sports pass quality rating (Horton et al., 2014). However, one challenge lies in delivering consistent, individualised models across team field sports that are dynamic in nature. For example, classification of soccer shots and passes showed a decline in model performance accuracy from a closed environment to a dynamic match setting (Schuldhaus et al., 2015). A method to overcome accuracy limitations in dynamic team field sports associated with solely using IMUs or vision may be to implement data fusion (Ó Conaire et al., 2010). Furthermore, vision and deep learning approaches have demonstrated the ability to track and classify team sport collective court activities and individual player specific movements in volleyball (Ibrahim et al., 2016), basketball (Ramanathan et al., 2015) and ice hockey (Tora et al., 2017). Accounting for methods from experimental set-up to model evaluation, previous reported models should be considered and adapted based on the current problem. Furthermore, the balance between model computational efficiency, results accuracy and complexity trade-offs calculations are an important factor.

In the present study, meta-analysis was considered however variability across developed model parameter reporting and evaluation methods did not allow for this to be undertaken. As this field expands and further methodological approaches are investigated, it would be practical to review analysis approaches both within and between sports. This review was delimited to machine and deep learning approaches to sport movement detection and recognition. However, statistical or parametric approaches not considered here such as discriminative functional analysis may also show efficacy for sport-specific movement recognition. However, as the field of machine learning is a rapidly developing area shown to produce superior results, a review encompassing all possible other methods may have complicated the reporting. Since sport-specific movements and their environments alter the data acquisition and analysis, the sports and movements reported in the present study provide an overview of the current field implementations.

5 Conclusions

This systematic review reported on the literature using machine and deep learning methods to automate sport-specific movement recognition. In addressing the research questions, both IMUs and computer vision have demonstrated capacity in improving the information gained from sport movement and skill recognition for performance analysis. A range of methods for model development were used across the reviewed studies producing varying results. Conventional machine learning algorithms such as Support Vector Machines and Neural Networks were most commonly implemented. Yet in those studies which applied deep learning algorithms such as Convolutional Neural Networks, these methods outperformed the machine learning algorithms in comparison. Typically, the models were evaluated using a leave-one-out cross validation method and reported model performances as a classification accuracy score. Intuitively, the adaptation of experimental set-up, data processing, and recognition methods used are best considered in relation to the characteristics of the sport and targeted movement(s). Consulting current models within or similar to the targeted sport and movement is of benefit to address bench mark model performances and identify areas for improvement. The application within the sporting domain of machine learning and automated sport analysis coding for consistent uniform usage appears currently a challenging prospect, considering the dynamic nature, equipment restrictions and varying environments arising in different sports.

Future work may look to adopt, adapt and expand on current models associated with a specific sports movement to work towards flexible models for mainstream analysis implementation. Investigation of deep learning methods in comparison to conventional machine learning algorithms would be of particular interest to evaluate if the trend of superior performances is beneficial for sport-specific movement recognition. Analysis as to whether IMUs and vision alone or together yield enhanced results in relation to a specific sport and its implementation efficiency would also be of value. In consideration of the reported study information, this review can assist future researchers in broadening investigative approaches for sports performance analysis as a potential to enhancing upon current methods.

Acknowledgements

Not applicable.

Disclosure statement

No potential conflict of interest was reported by the authors. However, the last author is co-author of three studies included in this systematic review.

Funding

The authors received no specific funding for this work.

Author ORCID

Sam Robertson 0000-0002-8330-0011

Alice Sweeting 0000-0002-9185-6773

Emily Cust 0000-0001-6927-6329

References

Adelsberger, R., & Tröster, G. (2013). Experts lift differently: Classification of weight-lifting athletes. In *2013 IEEE International Conference on Body Sensor Networks* (pp. 1–6). Cambridge, MA: Body Sensor Networks (BSN). https://doi.org/10.1109/BSN.2013.6575458

Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3D data: A review. *Pattern Recognition Letters*, *48*, 70–80. https://doi.org/10.1016/j.patrec.2014.04.011

Anand, A., Sharma, M., Srivastava, R., Kaligounder, L., & Prakash, D. (2017). Wearable motion sensor based analysis of swing sports. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 261–267). https://doi.org/10.1109/ICMLA.2017.0-149

Barris, S., & Button, C. (2008). A review of vision-based motion analysis in sport. *Sports Medicine*, *38*(12), 1025–1043. https://doi.org/10.2165/00007256-200838120-00006

Bengio, Y. (2013). Deep learning of representations: Looking forward. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7978 LNAI*, 1–37. https://doi.org/10.1007/978-3-642-39593-2_1

Bertasius, G., Park, H. S., Yu, S. X., & Shi, J. (2017). Am I a baller? Basketball performance assessment from first-person videos. *Proceedings of the IEEE International Conference on Computer Vision*, 2196–2204. https://doi.org/10.1109/ICCV.2017.239

Brock, H., & Ohgi, Y. (2017). Assessing motion style errors in ski jumping using inertial sensor devices. *IEEE Sensors Journal*, (99), 1–11. https://doi.org/10.1109/JSEN.2017.2699162

Brock, H., Ohgi, Y., & Lee, J. (2017). Learning to judge like a human: convolutional networks for classification of ski jumping errors. *Proceedings of the 2017 ACM International Symposium on Wearable Computers - ISWC ’17*, 106–113. https://doi.org/10.1145/3123021.3123038

Buckley, C., O’Reilly, M. A., Whelan, D., Vallely Farrell, A., Clark, L., Longo, V., … Caulfield, B. (2017). Binary classification of running fatigue using a single inertial measurement unit. In *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks* (pp. 197–201). IEEE. https://doi.org/10.1109/BSN.2017.7936040

Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, *46*(3), 1–33. https://doi.org/http://dx.doi.org/10.1145/2499621

Buthe, L., Blanke, U., Capkevics, H., & Tröster, G. (2016). A wearable sensing system for timing analysis in tennis. In *BSN 2016 - 13th Annual Body Sensor Networks Conference* (pp. 43–48). San Francisco, CA. https://doi.org/10.1109/BSN.2016.7516230

Bux, A., Angelov, P., & Habib, Z. (2017). Vision based human activity recognition: A review. In P. Angelov, A. Gegov, C. Jayne, & Q. Shen (Eds.), *Advances in Computational Intelligence Systems: Contributions Presented at the 16th UK Workshop on Computational Intelligence* (pp. 341–371). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-46562-3_23

Camomilla, V., Bergamini, E., Fantozzi, S., & Vannozzi, G. (2018). Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: a systematic review. *Sensors*, *18*(3), 873. https://doi.org/10.3390/s18030873

Chambers, R., Gabbett, T., Cole, M. H., & Beard, A. (2015). The use of wearable microsensors to quantify sport-specific movements. *Sports Medicine*, *45*(7), 1065–1081. https://doi.org/10.1007/s40279-015-0332-9

Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., Walsh, M., & O'Mathuna, C. (2011). Multi-sensor classification of tennis strokes. *Journal of IEEE Sensors*, 1437–1440.

Couceiro, M. S., Dias, G., Mendes, R., & Araújo, D. (2013). Accuracy of pattern detection methods in the performance of golf putting. *Journal of Motor Behavior*, *45*(1), 37–53. https://doi.org/10.1080/00222895.2012.740100

Díaz-Pereira, M. P., Gómez-Conde, I., Escalona, M., & Olivieri, D. N. (2014). Automatic recognition and scoring of olympic rhythmic gymnastic movements. *Human Movement Science*, *34*(1), 63–80. https://doi.org/10.1016/j.humov.2014.01.001

Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. M. P. (2010). Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, *14*(7), 645–662. https://doi.org/10.1007/s00779-010-0293-9

Fong, D. T.-P., & Chan, Y.-Y. (2010). The use of wearable inertial motion sensors in human lower limb biomechanics studies: A systematic review. *Sensors*, *10*(12), 11556–11565. https://doi.org/10.3390/s101211556

Gabbett, T., Jenkins, D., & Abernethy, B. (2012). Physical demands of professional rugby league training and competition using microtechnology. *Journal of Science and Medicine in Sport*, *15*, 80–86. https://doi.org/10.1016/j.jsams.2011.07.004

Gabbett, T., Jenkins, D. G., & Abernethy, B. (2011). Physical collisions and injury in professional rugby league match-play. *Journal of Science and Medicine in Sport*, *14*, 210–215. https://doi.org/10.1016/j.jsams.2011.01.002

Gastin, P. B., McLean, O. C., Breed, R. V., & Spittle, M. (2014). Tackle and impact detection in elite Australian football using wearable microsensor technology. *Journal of Sports Sciences*, *32*(10), 947–953. https://doi.org/10.1080/02640414.2013.868920

Gastin, P. B., McLean, O. C., Spittle, M., & Breed, R. V. (2013). Quantification of tackling demands in professional Australian football using integrated wearable athlete tracking technology. *Journal of Science and Medicine in Sport*, *16*(6), 589–593. https://doi.org/10.1016/j.jsams.2013.01.007

Gløersen, Ø., Myklebust, H., Hallén, J., & Federolf, P. (2018). Technique analysis in elite athletes using principal component analysis. *Journal of Sports Sciences*, *36*(2), 229–237. https://doi.org/10.1080/02640414.2017.1298826

Groh, B. H., Fleckenstein, M., & Eskofier, B. M. (2016). Wearable trick classification in freestyle snowboarding. In *13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)* (pp. 89–93). IEEE. https://doi.org/10.1109/BSN.2016.7516238

Groh, B. H., Fleckenstein, M., Kautz, T., & Eskofier, B. M. (2017). Classification and visualization of skateboard tricks using wearable sensors. *Pervasive and Mobile Computing*, *40*, 42–55. https://doi.org/10.1016/j.pmcj.2017.05.007

Groh, B. H., Kautz, T., & Schuldhaus, D. (2015). IMU-based trick classification in skateboarding. In *KDD Workshop on Large-Scale Sports Analytics*.

Hachaj, T., Ogiela, M. R., & Koptyra, K. (2015). Application of assistive computer vision methods to Oyama karate techniques recognition. *Symmetry*, *7*(4), 1670–1698. https://doi.org/10.3390/sym7041670

Hafer, J. F., & Boyer, K. A. (2017). Variability of segment coordination using a vector coding technique: reliability analysis for treadmill walking and running. *Gait and Posture*, *51*, 222–227. https://doi.org/10.1016/j.gaitpost.2016.11.004

Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. *Proceedings Of The International Joint Conference On Neural Networks*, *1*, 593–605. https://doi.org/10.1109/IJCNN.1989.118638

Hochreiter, S., & Schmidhuber, J. J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1–32. https://doi.org/10.1162/neco.1997.9.8.1735

Horton, M., Gudmundsson, J., Chawla, S., & Estephan, J. (2014). Classification of passes in football matches using spatiotemporal data. *ArXiv Preprint ArXiv:1407.5093*. https://doi.org/10.1145/3105576

Howe, S. T., Aughey, R. J., Hopkins, W. G., Stewart, A. M., & Cavanagh, B. P. (2017).

Quantifying important differences in athlete movement during collision-based team sports: Accelerometers outperform global positioning systems. In *2017 IEEE International Symposium on Inertial Sensors and Systems* (pp. 1–4). Kauai, HI, USA: IEEE. https://doi.org/10.1109/ISISS.2017.7935655

Hulin, B. T., Gabbett, T., Johnston, R. D., & Jenkins, D. G. (2017). Wearable microtechnology can accurately identify collision events during professional rugby league match-play. *Journal of Science and Medicine in Sport*, *20*(7), 638–642. https://doi.org/http://dx.doi.org/10.1016/j.jsams.2016.11.006

Ibrahim, M., Muralidharan, S., Deng, Z., Vahdat, A., & Mori, G. (2016). A Hierarchical Deep Temporal Model for Group Activity Recognition. *Cvpr*, 1971–1980. https://doi.org/10.1109/CVPR.2016.217

Jensen, U., Blank, P., Kugler, P., & Eskofier, B. M. (2016). Unobtrusive and energy-efficient swimming exercise tracking using on-node processing. *IEEE Sensors Journal*, *16*(10), 3972–3980. https://doi.org/10.1109/JSEN.2016.2530019

Jensen, U., Prade, F., & Eskofier, B. M. (2013). Classification of kinematic swimming data with emphasis on resource consumption. In *2013 IEEE International Conference on Body Sensor Networks, BSN 2013*. https://doi.org/10.1109/BSN.2013.6575501

Jensen, U., Schmidt, M., Hennig, M., Dassler, F. A., Jaitner, T., & Eskofier, B. M. (2015). An IMU-based mobile system for golf putt analysis. *Sports Engineering*, *18*(2), 123–133. https://doi.org/10.1007/s12283-015-0171-9

Ji, S., Yang, M., Yu, K., & Xu, W. (2013). 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 221–231. https://doi.org/10.1109/TPAMI.2012.59

Jiao, L., Wu, H., Bie, R., Umek, A., & Kos, A. (2018). Multi-sensor Golf Swing Classification Using Deep CNN. *Procedia Computer Science*, *129*, 59–65. https://doi.org/10.1016/j.procs.2018.03.046

Kapela, R., Świetlicka, A., Rybarczyk, A., Kolanowski, K., & O’Connor, N. E. (2015). Real-time event classification in field sport videos. *Signal Processing: Image Communication*, *35*, 35–45. https://doi.org/10.1016/j.image.2015.04.005

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014a). Large-scale video classification with convolutional neural networks. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference On*, 1725–1732. https://doi.org/10.1109/CVPR.2014.223

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014b). Large-scale video classification with convolutional nural networks. Retrieved December 18, 2017, from http://cs.stanford.edu/people/karpathy/deepvideo/

Kasiri-Bidhendi, S., Fookes, C., Morgan, S., Martin, D. T., & Sridharan, S. (2015). Combat sports analytics: Boxing punch classification using overhead depth imagery. In *2015 IEEE International Conference on Image Processing (ICIP)* (pp. 4545–4549). Quebec City, Canada: IEEE. https://doi.org/10.1109/ICIP.2015.7351667

Kasiri, S., Fookes, C., Sridharan, S., & Morgan, S. (2017). Fine-grained action recognition of boxing punches from depth imagery. *Computer Vision and Image Understanding*, *159*, 143–153. https://doi.org/10.1016/j.cviu.2017.04.007

Kautz, T. (2017). Acquisition, filtering and analysis of positional and inertial data in sports. *FAU Studies in Computer Science*, *2*.

Kautz, T., Groh, B. H., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B. M. (2017). Activity recognition in beach volleyball using a deep convolutional neural network. *Data Mining and Knowledge Discovery*, 1–28. https://doi.org/10.1007/s10618-017-0495-0

Ke, S. R., Thuc, H., Lee, Y. J., Hwang, J. N., Yoo, J. H., & Choi, K. H. (2013). A review on video-based human activity recognition. *Computers*, *2*, 88–131. https://doi.org/10.3390/computers2020088

Kelly, D., Coughlan, G. F., Green, B. S., & Caulfield, B. (2012). Automatic detection of collisions in elite level rugby union using a wearable sensing device. *Sports Engineering*, *15*(2), 81–92. Retrieved from https://0-link-springer-com.library.vu.edu.au/article/10.1007%2Fs12283-012-0088-5

Kobsar, D., Osis, S. T., Hettinga, B. A., & Ferber, R. (2014). Classification accuracy of a single tri-axial accelerometer for training background and experience level in runners. *Journal of Biomechanics*, *47*(10), 2508–2511. https://doi.org/10.1016/j.jbiomech.2014.04.017

Kos, M., & Kramberger, I. (2017). A wearable device and system for movement and biometric data

Acquisition for sports applications. *IEEE Access*, 1–1. https://doi.org/10.1109/ACCESS.2017.2675538

Kos, M., Ženko, J., Vlaj, D., & Kramberger, I. (2016). Tennis stroke detection and classification using miniature wearable IMU device. In *International Conference on Systems, Signals, and Image Processing*. https://doi.org/10.1109/IWSSIP.2016.7502764

Kotsiantis, S., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Informatica, 31,* 501–520. https://doi.org/10.1115/1.1559160

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances In Neural Information Processing Systems*, 1097–1105. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

Lai, D. T. H., Hetchl, M., Wei, X., Ball, K., & McLaughlin, P. (2011). On the difference in swing arm kinematics between low handicap golfers and non-golfers using wireless inertial sensors. *Procedia Engineering*, *13*, 219–225. https://doi.org/10.1016/j.proeng.2011.05.076

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

LeCun, Y., Bottou, L., Orr, G. B., & Müller, K. R. (1998). Efficient backprop. In *Neural Networks: Tricks of the Trade* (Vol. 1524, pp. 9–50).

LeCun, Y., Yoshua, B., & Geoffrey, H. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, J., Tian, Q., Zhang, G., Zheng, F., Lv, C., & Wang, J. (2018). Research on hybrid information recognition algorithm and quality of golf swing. *Computers and Electrical Engineering*, 1–13. https://doi.org/10.1016/j.compeleceng.2018.02.013

Liao, W. H., Liao, Z. X., & Liu, M. J. (2003). Swimming style classification from video sequences. In Kinmen (Ed.), *16th IPPR Conference on Computer Vision, Graphics and Image Processing* (pp. 226–233). ROC.

Lu, W. L., Okuma, K., & Little, J. J. (2009). Tracking and recognizing actions of multiple hockey players using the boosted particle filter. *Image and Vision Computing*, *27*(1–2), 189–205. https://doi.org/10.1016/j.imavis.2008.02.008

Magalhaes, F. A. de, Vannozzi, G., Gatta, G., & Fantozzi, S. (2015). Wearable inertial sensors in swimming motion analysis: A systematic review. *Journal of Sports Sciences*, *33*(7), 732–745. https://doi.org/10.1080/02640414.2014.962574

Mannini, A., & Sabatini, A. M. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, *10*(2), 1154–1175. https://doi.org/10.3390/s100201154

McNamara, D. J., Gabbett, T., Blanch, P., & Kelly, L. (2017). The relationship between wearable microtechnology device variables and cricket fast bowling intensity. *International Journal of Sports Physiology and Performance*, 1–20. https://doi.org/https://doi.org/10.1123/ijspp.2016-0540

McNamara, D. J., Gabbett, T., Chapman, P., Naughton, G., & Farhart, P. (2015). The validity of microsensors to automatically detect bowling events and counts in cricket fast bowlers. *International Journal of Sports Physiology and Performance*, *10*(1), 71–75. https://doi.org/10.1123/ijspp.2014-0062

Minnen, D., Westeyn, T. L., Starner, T., Ward, J. a, & Lukowicz, P. (2006). Performance metrics and evaluation issues for continuous activity recognition. In *Proc. Int. Workshop on Performance Metrics for Intelligent Systems* (pp. 141–148). https://doi.org/10.1145/1889681.1889687

Mitchell, E., Monaghan, D., & O'Connor, N. E. (2013). Classification of sporting activities using smartphone accelerometers. *Sensors (Basel, Switzerland)*, *13*(4), 5317–5337. https://doi.org/10.3390/s130405317

Mohammed, M., Khan, M., & Bashier, E. (2016). *Machine Learning: Algorithms and Applications*. Milton: CRC Press.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Montoliu, R., Martín-Félez, R., Torres-Sospedra, O., & Martínez-Usó, A. (2015). Team activity recognition in Association football using a bag-of-words-based method. *Human Movement Science*, *41*, 165–178. https://doi.org/10.1016/j.humov.2015.03.007

Mooney, R., Corley, G., Godfrey, A., Quinlan, L. R., & ÓLaighin, G. (2015). Inertial sensor technology for elite swimming performance analysis: A systematic review. *Sensors*, *16*(1),

18. https://doi.org/10.3390/s16010018

Nibali, A., He, Z., Morgan, S., & Greenwood, D. (2017). Extraction and classification of diving clips from continuous video footage. *ArXiv, pre-print*. Retrieved from https://arxiv.org/pdf/1705.09003.pdf

O'Reilly, M. A., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017a). Classification of lunge biomechanics with multiple and individual inertial measurement units. *Sports Biomechanics*, *16*(3), 342–360. https://doi.org/10.1080/14763141.2017.1314544

O'Reilly, M. A., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017b). Technology in strength and conditioning tracking lower-limb exercises with wearable sensors. *Journal of Strength and Conditioning Research*, *31*(6), 1726–1736.

O'Reilly, M., Caulfield, B., Ward, T., Johnston, W., & Doherty, C. (2018). Wearable Inertial Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review. *Sports Medicine*. https://doi.org/10.1007/s40279-018-0878-4

O'Reilly, M., Whelan, D., Chanialidis, C., Friel, N., Delahunt, E., Ward, T., & Caulfield, B. (2015). Evaluating squat performance with a single inertial measurement unit. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks*. IEEE. https://doi.org/10.1109/BSN.2015.7299380

O'Reilly, M., Whelan, D. F., Ward, T. E., Delahunt, E., & Caulfield, B. (2017). Classification of deadlift biomechanics with wearable inertial measurement units. *Journal of Biomechanics, 58,* 155–161. https://doi.org/10.1080/14763141.2017.1314544

Ó Conaire, C., Connaghan, D., Kelly, P., O'Connor, N. E., Gaffney, M., & Buckley, J. (2010). Combining inertial and visual sensing for human action recognition in tennis. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams* (pp. 51–56). ACM. https://doi.org/10.1145/1877868.1877882

Pernek, I., Kurillo, G., Stiglic, G., & Bajcsy, R. (2015). Recognizing the intensity of strength training exercises with wearable sensors. *Journal of Biomedical Informatics*, *58*, 145–155. https://doi.org/10.1016/j.jbi.2015.09.020

Plötz, T., Hammerla, N. Y., & Olivier, P. (2011). Feature learning for activity recognition in ubiquitous computing. *International Joint Conference on Artificial Intelligence (IJCAI),* 1729.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing*, *28*(6), 976–990. https://doi.org/10.1016/j.imavis.2009.11.014

Preece, S. J., Goulermas, J. Y., Kenney, L., & Howard, D. (2009). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering*, *56*(3), 871–879. https://doi.org/10.1109/TBME.2008.2006190

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors: A review of classification techniques. *Physiological Measurement*, *30*(4), R1–R33. https://doi.org/10.1088/0967-3334/30/4/R01

Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, *42*(3), 203–231. https://doi.org/10.1023/A:1007601015854

Qaisar, S., Imtiaz, S., Glazier, P., Farooq, F., Jamal, A., Iqbal, W., & Lee, S. (2013). A method for cricket bowling action classification and analysis using a system of inertial sensors. In *International Conference on Computational Science and its Applications* (pp. 396–412). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39649-6

Ramanathan, V., Huang, J., Abu-El-Haija, S., Gorban, A., Murphy, K., & Fei-Fei, L. (2015). Detecting events and key actors in multi-person videos. https://doi.org/10.1109/CVPR.2016.332

Rassem, A., El-Beltagy, M., & Saleh, M. (2017). Cross-country skiing gears classification using deep learning. *ArXiv Preprint ArXiv:1706.08924*. Retrieved from https://arxiv.org/pdf/1706.08924v1.pdf

Ravi, D., Wong, C., Lo, B., & Yang, G.-Z. (2016). A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE Journal of Biomedical and Health Informatics*, *21*(1), 1–1. https://doi.org/10.1109/JBHI.2016.2633287

Reily, B., Zhang, H., & Hoff, W. (2017). Real-time gymnast detection and performance analysis with a portable 3D camera. *Computer Vision and Image Understanding*, *159*, 154–163. https://doi.org/10.1016/j.cviu.2016.11.006

Rindal, O. M. H., Seeberg, T. M., Tjønnås, J., Haugnes, P., & Sandbakk, Ø. (2018). Automatic classification of sub-techniques in classical cross-country skiing using a machine learning algorithm on micro-sensor data. *Sensors (Switzerland), 18*(1), 75. https://doi.org/10.3390/s18010075

Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications*, *59*, 235–244. https://doi.org/10.1016/j.eswa.2016.04.032

Saba, T., & Altameem, A. (2013). Analysis of vision based systems to detect real time goal events in soccer videos. *Applied Artificial Intelligence*, *27*(7), 656–667. https://doi.org/10.1080/08839514.2013.787779

Salman, M., Qaisar, S., & Qamar, A. M. (2017). Classification and legality analysis of bowling action in the game of cricket. *Data Mining and Knowledge Discovery*, *31*(6), 1706–1734. https://doi.org/10.1007/s10618-017-0511-4

Schuldhaus, D., Zwick, C., Körger, H., Dorschky, E., Kirk, R., & Eskofier, B. M. (2015). Inertial sensor-based approach for shot/ pass classification during a soccer match. In *Proc. 21st ACM KDD Workshop on Large-Scale Sports Analytics* (pp. 1–4). Sydney, Australia.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2008). RUSBoost: Improving classification performance when training data is skewed. In *9th International Conference on Pattern Recognition* (pp. 1–4). https://doi.org/10.1109/ICPR.2008.4761297

Shah, H., Chokalingam, P., Paluri, B., & Pradeep, N. (2007). Automated stroke classification in tennis. *Image Analysis and Recognition*, 1128–1137.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: from theory to algorithms*. New York, USA: Cambridge University Press.

Sharma, M., Srivastava, R., Anand, A., Prakash, D., & Kaligounder, L. (2017). Wearable motion sensor based phasic analysis of tennis serve for performance feedback. In *2017 IEEE International Conference on Acoustics, Speech and Siginal Processing* (pp. 5945–5949). New Orleans, LA: IEEE.

Sprager, S., & Juric, M. B. (2015). *Inertial sensor-based gait recognition: A review*. *Sensors (Switzerland)* (Vol. 15). https://doi.org/10.3390/s150922089

Srivastava, R., Patwari, A., Kumar, S., Mishra, G., Kaligounder, L., & Sinha, P. (2015). Efficient characterization of tennis shots and game analysis using wearable sensors data. In *2015 IEEE Sensors- Proceedings* (pp. 1–4). Busan. https://doi.org/10.1109/ICSENS.2015.7370311

Stein, M., Janetzko, H., Lamprecht, A., Breitkreutz, T., Zimmermann, P., Goldlücke, B., … Keim, D. A. (2018). Bring it to the pitch: combining video and movement data to enhance team sport analysis. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 13–22. https://doi.org/10.1109/TVCG.2017.2745181

Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. (2017). Efficient processing of deep neural networks: A tutorial and survey. *IEEE*, *105*(2), 2295–2329. Retrieved from http://arxiv.org/abs/1703.09039

Thomas, G., Gade, R., Moeslund, T. B., Carr, P., & Hilton, A. (2017). Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, *159*, 3–18. https://doi.org/10.1016/j.cviu.2017.04.011

Titterton, D. H., & Weston, J. L. (2009). *Strapdown inertial navigation technology* (2nd ed.). Reston, VA: AIAA.

Tora, M. R., Chen, J., & Little, J. J. (2017). Classification of puck possession events in ice hockey. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 147–154). https://doi.org/10.1109/CVPRW.2017.24

Victor, B., He, Z., Morgan, S., & Miniutti, D. (2017). Continuous video to simple signals for swimming stroke detection with convolutional neural networks. *ArXiv Preprint ArXiv:1705.09894*. https://doi.org/10.1111/j.1467-8330.1974.tb00606.x

Wagner, D., Kalischewski, K., Velten, J., & Kummert, A. (2017). Activity recognition using inertial sensors and a 2-D convolutional neural network. In IEEE (Ed.), *2017 10th International Workshop on Multidimensional (nD) Systems (nDS)* (pp. 1–6). https://doi.org/10.1109/NDS.2017.8070615

Wagner, J. F. (2018). About motion measurement in sports based on gyroscopes and accelerometers - an engineering point of view. *Gyroscopy and Navigation*, *9*(1), 1–18. https://doi.org/10.1134/S2075108718010091

Ward, J. A., Lukowicz, P., & Gellersen, H.-W. (2011). Performance metrics for activity recognition. In *ACM Trans. on Intelligent Systems and Technology* (Vol. 2, pp. 111–132).

Whiteside, D., Cant, O., Connolly, M., & Reid, M. (2017). Monitoring hitting load in tennis using inertial sensors and machine learning. *International Journal of Sports Physiology and*

Performance, 1–20. https://doi.org/https://doi.org/10.1123/ijspp.2016-0683

Wixted, A., Billing, D. C., & James, D. A. (2010). Validation of trunk mounted inertial sensors for analysing running biomechanics under field conditions, using synchronously collected foot contact data. *Sports Engineering*, *12*(4), 207–212. https://doi.org/10.1007/s12283-010-0043-2

Wixted, A., Portus, M., Spratford, W., & James, D. A. (2011). Detection of throwing in cricket using wearable sensors. *Sports Technology*, *4*(3–4), 134–140. https://doi.org/10.1080/19346182.2012.725409

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, *8*(7), 1341–1390. https://doi.org/10.1162/neco.1996.8.7.1391

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimisation. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82. https://doi.org/10.1023/A:1021251113462

Wundersitz, D. W., Gastin, P. B., Richter, C., Robertson, S., & Netto, K. J. (2015). Validity of a trunk-mounted accelerometer to assess peak accelerations during walking, jogging and running. *European Journal of Sport Science*, *15*(5), 382–390. https://doi.org/10.1080/17461391.2014.955131

Wundersitz, D. W., Gastin, P. B., Robertson, S., Davey, P. C., & Netto, K. J. (2015). Validation of a trunk-mounted accelerometer to measure peak impacts during team sport movements. *International Journal of Sports Medicine*, *36*(9), 742–746. https://doi.org/10.1055/s-0035-1547265

Wundersitz, D. W., Josman, C., Gupta, R., Netto, K. J., Gastin, P. B., & Robertson, S. (2015). Classification of team sport activities using a single wearable tracking device. *Journal of Biomechanics*, *48*(15), 3975–3981. https://doi.org/10.1016/j.jbiomech.2015.09.015

Yang, C. C., & Hsu, Y. L. (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, *10*(8), 7772–7788. https://doi.org/10.3390/s100807772

Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., & Shonali, K. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of the 24th International Conference on Artificial Intelligence* (pp. 3995–4001).

Yao, B., & Fei-Fei, L. (2010). Modeling mutual context of object and human pose in human-object interaction activities. In *Computer Vision and Pattern Recognition* (pp. 17–24). IEEE.

Young, C., & Reinkensmeyer, D. J. (2014). Judging complex movement performances for excellence: a principal components analysis-based technique applied to competitive diving. *Human Movement Science*, *36*, 107–122. https://doi.org/10.1016/j.humov.2014.05.009

Yu, G., Jang, Y. J., Kim, J., Kim, J. H., Kim, H. Y., Kim, K., & Panday, S. B. (2016). Potential of IMU sensors in performance analysis of professional alpine skiers. *Sensors (Switzerland)*, *16*(4), 1–21. https://doi.org/10.3390/s16040463

Zebin, T., Scully, P. J., & Ozanyan, K. B. (2016). Human Activity Recognition with Inertial Sensors Using a Deep Learning Approach. *Proc. of IEEE Sensors 2016*, (1), 1–3. https://doi.org/10.1109/ICSENS.2016.7808590

Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014). Convolutional neural networks for human activity recognition using mobile sensors. In *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services* (pp. 197–205). https://doi.org/10.4108/icst.mobicase.2014.257786

Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., & Li, Z. (2017). A review on human activity recognition using vision-based method. *Journal of Healthcare Engineering*, *2017*, 1–31. https://doi.org/10.1155/2017/3090343

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). Time series classification using multi-channels deep convolutional neural networks. In *International Conference on Web-Age Information Management* (pp. 298–310). Springer. https://doi.org/10.1007/978-3-319-08010-9_33

Zhu, G., Xu, C., Gao, W., & Huang, Q. (2006). Action recognition in broadcast tennis video. *Computer Vision in Human-Computer Interaction*, 89–98. https://doi.org/10.1007/11754336_9

Ziaeefard, M., & Bergevin, R. (2015). Semantic human activity recognition: A literature review. *Pattern Recognition*, *48*(8), 2329–2345. https://doi.org/10.1016/j.patcog.2015.03.006